Author:
Griffith, Gareth

Title:
How can we best understand mental health in the UK?

Critically and empirically investigating the measurement, and socio-demographic and geographical predictors of UK mental health.

How can we best understand Mental Health in the UK?

Critically and Empirically Investigating the Measurement, and Socio-Demographic and Geographical Predictors of UK Mental Health.

Gareth Griffith

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of PhD in Geography / Advanced Quantitative Methods in the Faculty of Social Sciences and Law, School of Geographical Science, April 2018

Word Count: 67,784

ABSTRACT

Mental Health is an issue of increasingly accepted importance in the UK, both in policy and in the public sphere. It becomes clear under scrutiny, however, that any attempt to characterise and prioritise "at risk" groups risks making conclusions entirely sensitive to the definition of mental health chosen by the researcher. This thesis examines the definition and distribution of the mental health conditions which have been the backdrop for this increase in awareness. This involves combining innovative quantitative methodologies in order to develop a comprehensive understanding of UK mental health from 2009-2016. The thesis is divided into two halves.

The first half addresses the definition of mental health. Mental health measurement can be crudely categorised into theories which characterise mental health via absence of negative elements and presence of positive elements. Unpacking what is truly captured by the most widely used population screening metrics for each is undertaken using Exploratory Structural Equation Modelling (ESEM). Solutions for both metrics are presented and demonstrated to be both substantively and empirically superior to traditional summed interpretation.

The second half of the thesis incorporates these newly suggested interpretations into an investigation of the distribution of mental health across the UK using advanced multilevel modelling techniques. Firstly, a demonstration of what could be known given the summed interpretation of the metrics is given. This demonstrates the differences between the geographical and demographic determinants of the metrics. Finally, the constructs developed from the ESEM analysis become responses in a longitudinal investigation of chronicity in mental health.

This presents the first comprehensive overview of UK mental health incorporating both ESEM and Multilevel Modelling. It represents clear evidence of the synergy of the two methodologies. Moreover, the empirical findings of the thesis are relevant and important for central and local governments in developing geographically-sensitive mental health policy.

ACKNOWLEDGEMENTS

Firstly, I would like to extend enormous gratitude to Kelvyn Jones for being my lead supervisor throughout this PhD. His knowledge, insight, support and friendship have been invaluable, and this project certainly would not have been possible without his excellent presence. Thanks also to David Manley, George Leckie and Dewi Owen who have all provided excellent advice and guidance during the course of this work. This is also extended to the wider Spatial Modelling Group at Bristol, and Centre for Multilevel Modelling, who have helped prepare presentations and provide guidance in a friendly, collaborative environment. Thanks also to the ESRC for funding me to carry out this research, and the UK Data Service and ISER for providing access to Understanding Society Data.

I would also like to extend my thanks to friends and family for their support, and for not entirely shunning me during my extended absence whilst writing up. Particular thanks to my parents and sisters, and Gwilym Owen, Dewi Owen, Tim Morris, Pete Crowder, Becca Toop, Veronica Wignall and Josh Keyworth for helping with proof-reading, and keeping me sane. Special thanks also to my grandfather, Robert Yon, to whom I dedicate this thesis. Thank you for enabling all this by making the trip to the UK and being a continuing source of inspiration and understanding, your unswerving humility forever belies your true impact.

Finally, thanks to the Postgraduates and staff in the Geographical Sciences Department for creating one of the most interesting and collaborative cross-discipline research environments I have ever had the fortune to be a part of. I sincerely hope my future lunchtime discussions are as interesting, informed and entertaining as those I had in my years in Browns.

AUTHORS DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, other, is indicated as such. Any views expressed in the thesis are those of the author.

SIGNED:…………………………………………………… DATE:………………..

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introducing the Mental Health of the UK

1.1 The growing acceptance of the importance of mental health

1.1.1 Health Burden

The topic of mental health has received a remarkable increase in global attention in recent years. A large body of research seems to suggest that there is a current global trend towards increasing numbers of individuals with depressive disorders (Levinson *et al.*, 2010; Lépine and Briley, 2011; World Health Organization, 2013). This overall increase in mental health burden is also evident in the UK, where a similar trend has been identified for prescriptions, deaths by suicide and population screening metrics (Barr *et al.*, 2012; Katikireddi, Niedzwiedz and Popham, 2012; Spence *et al.*, 2014; Barr, Kinderman and Whitehead, 2015; Collishaw, 2015). This has provoked an increased interest in the topic of mental health in the UK. This was particularly noticeable in the recent elections, with 2015 being the first year that all major UK parties included specific mental health goals in their manifestos. The UK Centre for Mental Health reported in 2015 that mental health problems account for the largest proportion of the total disease burden in the UK at 28%, compared with 16% apiece for cancer and heart disease (Ferrari *et al.*, 2013). It has been suggested that recent economic austerity and budget tightening have fostered growing job insecurity and unemployment in the country which, coupled with increased awareness has contributed to negative mental health trends since the 2008 global financial crisis. A particularly stark finding comes from Barr et al. (2012), who find the recent recession led to 1000 excess suicides in the years 2008-2010 relative to previous trends. Similarly, Stuckler et al. (2009) found that across Europe, there has been an increase of 0.79% in suicides for every 1% increase in unemployment for those under the age of 65.

1.1.2 Economic Burden

This well-documented increase in mental illness also has a significant associated financial burden. The 2013 report of the Chief Medical Officer found that mental illness costs the UK economy £70-100bn per annum, 4.5% of the total Gross Domestic Product (Davies, 2013). Furthermore, the same report stated that approximately 75% of individuals with mental illness receive absolutely no treatment. There are clear causal pathways between mental illness and lowered social and economic productivity, with poor mental health both predicting and being a result of unemployment (Paul and Moser, 2009; Stuckler *et al.*, 2009; Flint, Bartley, *et al.*, 2013). Thus, the link between economic prosperity and mental health is important not only because it may predict economic struggle, but that it is also strongly impacted by existing hardship.

1.1.3 Promotion of Awareness

It is well acknowledged that there is a disparity between the investment and research into mental health and similarly burdening physical ailments (e.g. Tomlinson et al., 2009). Globally, awareness of the burden of mental disorders has also increased in recent years. Major unipolar depressive disorder was declared the second leading cause of disability worldwide as measured by disability-adjusted life-years (DALYs) for the year 2000 by the World Health Organisation (WHO, 2001). More recently, Whiteford et al. (2013) found in the Global Burden of Disease Study from 2010 that mental and substance use disorders account for about 7.4% of total worldwide disease burden, and over the next 20 years mental health conditions could account for the loss of 16.1 trillion US$ worldwide (Bloom *et al.*, 2011). Moreover, it has been suggested that markers of wellbeing and positive mental health should be considered a better representation of a nation’s development than more traditional measures of economic productivity (Stiglitz, Sen and Fitoussi, 2009). However, with the notable exception of Bhutan who have developed a Gross National Happiness Index (GNH) as an indicator of progress (Braun, 2009), this recommendation has yet to make it into wider governmental legislative policy.

The disparity between physical and mental health promotion and awareness persists despite recent additional large scale funding for mental health initiatives, for example the promotion of the “No Health without Mental Health” strategy in the UK in 2011 (HM Government, 2012). The disparity in awareness and funding undoubtedly contributes to the large scale under-diagnosis of mental health conditions - with 70% of individuals who suffer globally from mental illness never receiving treatment (Henderson, Evans-Lacko and Thornicroft, 2013), highlighting the critical need for greater understanding of the scale and nature of the issue.

The increased awareness of mental health as an issue of concern for today’s society has highlighted the need for metrics to capture improvement. There needs to be a demonstrable benefit of a policy intervention, otherwise how can a government be sure that it is worth continued investment or promotion? In the case of mental health, as will become clear, capturing change requires measurement that is far from agreed upon.

1.2 Chapter Outline

This introductory chapter starts by giving a brief overview of some of the dominant themes of this thesis; namely ‘unpacking’ the measurement of mental health and investigating differential demographic risk. These themes are formalised into research questions which become the basis of each subsequent chapter. It then gives a brief overview of methodological advances and developments that will be used in this thesis to better understand these notions. Finally, it concludes with an overview of the subsequent chapters, summarising the motive and methods for each section.

1.3 Measurement becomes Understanding

1.3.1 Understanding Mental Health – From Clinical Outcomes to Population Screening

Despite these large-scale acknowledgements of the importance of understanding mental health and mood disorders, there is relatively little consensus on the best methods for measuring these to assess prevalence and aetiology. There is a longstanding literature on the difficulty that lies in diagnosing mental illness due to the inherent lack of easily quantifiable material symptoms on which to base diagnoses. Specifically in mood disorders the lack of easily recognisable, tangible, physical symptoms are suggested to contribute to the very high estimated incidences of non-diagnosis (Verheij, 1996; Davies, 2013). Due to this, a large body of early mental epidemiological research focused on more easily evidenced psychoses, such as schizophrenia or substance abuse, due to their relative ease of quantification via clinical outcomes such as hospital admissions (e.g. Giggs, 1973; Kilpatrick et al., 2000; Silver, Mulvey, & Swanson, 2002). These early studies necessarily underestimated the full burden of mental illness as ultimately the overwhelming majority of individuals suffering from mental illness never receive treatment (Wittchen and Jacobi, 2005; Thornicroft, 2007; Clement *et al.*, 2015). As this literature emerged demonstrating the burden of Common Mood Disorders (CMDs) such as depression (e.g. World Health Organization, 2001, 2013; Weich *et al.*, 2004), mental health research broadened to try and incorporate these new perspectives.

The increased appetite for characterising mood disorders required a fundamentally different approach than analyses of psychosis. Where mental health could previously be characterised by the “absence of illness” in terms of psychosis, mood disorders started to introduce the notion of a continuum of mental health (Keyes, 2002; Westerhof and Keyes, 2010). Attempts to define and measure this underlying continuum have resulted in the development of a multitude of measurement questionnaires aiming to capture this underlying psychological distress (e.g. Goldberg, 1972; Tennant et al., 2007). It is this relatively new area that will be explored in depth in this thesis, specifically the measurement of depression and anxiety via a multi-item questionnaire in the general population, and the subsequent analysis of their distribution and socio-demographic determinants across the UK in the years following the 2008 global economic recession.

In advancing our understanding of mental health, there are several key themes which present themselves. These can be broadly defined into three domains:

- Issues of *definition;* what is being referred to when people speak about mental health;
- Issues of *awareness*; which broadly go hand-in-hand with issues of stigma and understanding;
- Issues of *risk*; concerning where and whom are those who are most vulnerable and how they can be helped.

This chapter outlines the approach to these issues that will be undertaken in this thesis. It first considers what is commonly understood when referring to mental health, and gives a brief overview of how it is currently measured. Secondly it gives a brief outline of the social and demographic characteristics which have been shown to be associated with mental health, and discusses the effect of geographical as well as social context on mental health. The chapter concludes in an outline of the final thesis, summarising the content of each of the coming chapters.

1.3.2 Understanding Mental Health – From Mental Illness to Mental Well-being:

The fundamental issue with measuring depressive symptoms is that no direct measurement is universally accepted. One cannot ask an individual how depressed they are and receive an objective and comparable response that is consistent across individuals, thus findings must be considered as potentially measurement-specific. Mental health is inherently subjective, something which is suggested to have contributed to the stigma surrounding it. However, although individuals live in objectively categorised worlds, their subjective experience of this world is what dictates their responses, and therefore our understanding of their mental health (Keyes, Shmotkin and Ryff, 2002). There is a great deal of early work on mental health which essentially debates the validity of subjective measures (Bradburn, 1964; Phillips, Andrews and Withey, 1978; Massé *et al.*, 1998) and set in motion the shift in understanding of mental health from categorised objective threshold to subjective spectrum. The emergence of mental health as a spectrum led to the increasing conceptualisation of an underlying latent (unmeasurable) trait of mental health, which can only be observed indirectly by asking proxy questions relating to it (Greenspoon and Saklofske, 2001; Keyes, 2002; Brodersen *et al.*, 2007). The move toward mood-disorders from psychoses in mental health literature has been gradual but understandable given this more holistic understanding.

Moving from more easily evidenced psychoses to including and addressing mood disorders has undoubtedly been a positive step for public health perspectives as it recognizes a much broader range of harmful conditions. This is evident in the UK with the publishing of a range of reports on mental health policy in recent years which use this framework to robustly prioritise and justify efforts for mental health improvement (e.g. Department of Health, 2009; The Scottish Government, 2009; Department of Health and Health, 2011; Davies, 2013). Despite this acknowledged importance however, the definition of what constitutes individual mental health remains the subject of intense debate. More traditional views of mental health have tended to focus on the objective of “absence of illness” – inferring mental wellness from this (Westerhof and Keyes, 2010). However, in recent years the notion of “wellbeing” has re-emerged, having first appeared in the 1950s as part of a movement to measure social change and improve social policy (Land, 1975).

Wellbeing advocates’ primary focus has been on broadening the understanding of “mental health” to include positive mental health functioning, of mental wellbeing as more than simply the absence of illness (World Health Organization, 2004; Slade, 2010; Levin *et al.*, 2012). Furthermore, positive mental health has been recognised as having significant implications for both physical health and broader social outcomes (e.g. Huppert, Baylis and Keverne, 2005; Howell, Kern and Lyubomirsky, 2007; Slade, 2010; Stiglitz, Sen and Fitoussi, 2010; Diener and Chan, 2011; Knapp, McDaid and Parsonage, 2011; Boehm and Kubzansky, 2012; Keyes and Simoes, 2012).

1.3.3 Understanding Mental Health – Positive Mental Health

Mental wellbeing is typically viewed as having two underlying dimensions, the hedonic and the eudemonic (Keyes, 2005; Westerhof and Keyes, 2010). Hedonic wellbeing refers to the emotional state of the individual, and is chiefly concerned with the pursuit of pleasure and happiness and the avoidance of pain (Chanfreau *et al.*, 2013). Eudemonic wellbeing goes beyond the emotional to the functioning of the individual, relating to social relationships and perceptions of their undertakings as worthwhile or meaningful (Ryan and Deci, 2001). There is, however, acknowledged understanding that wellbeing can encompass far more than these two domains (Connell, O’Cathain and Brazier, 2014).

The emergence of these two notions as separate entities, that of mental illness and that of mental wellbeing, has been positive for the development of the mental health literature. However, there is still considerable debate on what they represent. Despite the widely acknowledged differences in the constructs in psychological literature, they are often used interchangeably in policy documents (e.g. Helmchen & Lo Sasso, 2010). It is common for policy to readily translate findings from mental illness literature to recommendations for wellbeing, with absence of mental illness used to suggest high wellbeing, missing the nuances in the construct. This conflation of terms serves to undermine our understanding of the determinants of each, given that they have been shown to differ despite the terms seeming interrelatedness (e.g. Hu, Stewart-Brown, Twigg, & Weich, 2007; Stewart-Brown, Samaraweera, Taggart, Kandala, & Stranges, 2015). Set against this backdrop, there is a clear need for greater clarity in terminology, greater understanding of what each construct means, and a greater understanding of what is truly being captured in instruments developed to capture them.

1.3.4 Measuring becomes Understanding – Moving Forward

This increased acknowledgement of the importance of both positive and negative mental health and mental health measurement is evidenced in national research priorities. Most recently this was evidenced in the decision by the UK Office for National Statistics (ONS) to include four "Subjective Well-being" measures in the Annual Population Survey for the first time in 2011 (Waldron, 2010, Office for National Statistics, 2016). This was in direct response to a seminal paper by the Stiglitz Commission in 2009 which stressed the narrowness of traditional economic development measures such as Gross Domestic Product, and advocated the inclusion of wellbeing measures as an indicator of development. The increasing adoption of mental health as a country-level priority has taken place alongside and in response to the progress being made beyond the archetypal view that nothing more can be done for the mental health of individuals without a diagnosed mental illness.

1.4 Measuring Mental Health – Clinical Outcomes to Screening Questionnaires:

As suggested above, clinical outcomes are problematic for mental health research as they vastly under-represent the full burden of mood disorders. This is not a fault of the researchers using them, it is simply a product of under-diagnosis. It is estimated that the majority of individuals suffering from mental illness in Europe and the United States never receive treatment (Verheij, 1996; Wittchen and Jacobi, 2005; Thornicroft, 2007; Davies, 2014; Clement *et al.*, 2015). This is suggested to be an even higher proportion in lower or middle-income settings where provision is poorer (Wang *et al.*, 2007).

Moving past clinical outcomes as mental health measures requires a method which can be distributed to a population. Questionnaires are the most commonly used method for monitoring mental health on a large scale (Bowling, 2005). This is in part because they can be distributed widely, and repeatedly without having to constantly reinterpret the information. This last point has critical importance as, assuming the individual level understanding of each item does not change rapidly between years, it allows the measures to be interpreted over time. Without this temporal consistency, it is not possible to assume that the same thing is being captured at each occasion meaning researchers would be unable to characterise an individual's trajectory, let alone identify worsening and improving of the mental health of the population.

Mental health questionnaires can be crudely categorised into two groups, those that measure mental health via the presence of illness symptoms or lack thereof, and those that measure mental health via the presence of positive psychological phenomena. Of course, there is some crossover between the two categories, with many questionnaires containing elements of both. Moreover, questionnaire content tends to differ depending on when it was conceived, earlier efforts tending to be focus more on mental illness. More recent measures tend to focus more on positive mental health measurement, reflecting the broader understanding and appreciation of mental health at that time. One of each of these types of mental health measures are used and evaluated in this thesis: the 12-Item General Health Questionnaire and the Short Warwick-Edinburgh Mental Well-Being Scale. From a contemporary epidemiological perspective, the ideal standard would be a positive mental health measure, which could be distributed as a routine monitoring questionnaire, which would still retain the capacity to detect morbidity in those at the very low end of the spectrum.

1.4.1 The 12-Item General Health Questionnaire (GHQ-12)

Of these traditional 'illness' style questionnaires there is arguably no more widely adopted example than the 12-item General Health Questionnaire (GHQ-12) (Werneke *et al.*, 2000; Ye, 2009). It was initially designed in 1972 as a 60-Item questionnaire which aimed to capture both inability to carry out normal functioning, and emergence of new distressing phenomena in the respondent (Goldberg & Williams, 1988). Goldberg subsequently suggested the shorter 12-item version which was demonstrated to retain the capacity to assess and detect psychiatric morbidity (meaning psychological illness) in both community and non-psychiatric settings (Goldberg and Hillier, 1979). The brevity of this shorter version led to its rapid and widespread adoption as an outcome of interest in quantitative analyses of mental health. As of 2006, the instrument had been translated into 38 different languages, a strong testament to its perceived cross-cultural reliability and validity (Jackson, 2006)[1]. Whilst there is some evidence that the GHQ-12 does capture positive mental health elements (Hu *et al.*, 2007), its initial focus on diagnostic capacity has led to other wellbeing measures being developed in its stead as a supplement or indeed as a replacement.

1.4.2 The Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS)

With respect to wellbeing measures there are also several options, however among the most widely distributed in the UK is the 7-Item Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS). The original 14-Item scale was developed in 2007 by Tennant et al., and was designed to capture both hedonic and eudemonic elements of positive wellbeing amongst the population. It was subsequently shortened, similarly to the GHQ-12, for ease of use in a clinical setting (Stewart-Brown *et al.*, 2009). Again, similarly to the GHQ-12, the SWEMWBS has been extensively validated across cultures and languages (e.g. Bartram, Sinclair, & Baldwin, 2013; S. L. Stewart-Brown et al., 2011; Taggart et al., 2013).

The first half of this thesis is concerned with measurement and will therefore attempt to disentangle and critically evaluate what is being captured by each of these mental health measures when applied to the general population, and critically evaluate the assumptions inherent in their usage. It will then go on to evaluate the consistency of what is being measured across each scale, using a dataset an order of magnitude larger than that which has been used in the WEMWBS and GHQ original validation studies (Goldberg *et al.*, 1997; Tennant *et al.*, 2007). The second half of this thesis is concerned with relative risk as captured by the measures

---

[1] For further details on country and context-specific validation studies see Chapter 2.

outlined in the first half. It investigates how this relative risk can be stratified or investigated relative to social characteristics.

1.5 *Mental Health Inequality:*

The burden of mental distress is not felt uniformly across the UK. Policy interventions for mental health need to be focused in areas and demographic groups which are most burdened by mental distress and would therefore benefit most from the intervention. Understanding which demographic groups and geographical areas are at greater risk of poor mental health is therefore, given limited resources, of critical importance in improving mental health nationwide.

1.5.1 The Presence of Social Gradients in Mental Health

Demographic patterning exists in mental health and is widely reported in the literature. A multitude of these different characteristics have been shown to be associated with individual level mental health. However, these relationships have been shown to be sensitive to both measurement and methodology. Determinants differ for positive and negative mental health, and more realistically nuanced conclusions are only feasible in more complex modelling frameworks (Huppert and Whittington, 2003; Propper *et al.*, 2005; Keyes and Simoes, 2012). As such, it is important that the most realistic and nuanced models are used. This thesis will re-evaluate claims made in previous literature with respect to these predictors using a large-scale UK-specific dataset and advanced methodological techniques. A brief outline of some of the socio-demographic associations investigated in this thesis is given below, which are explored in greater detail in Chapter 4.

Gender has been shown to be strongly impactful on mental health. Females have been consistently found to report higher levels of psychological distress (e.g. McLean, Asnaani, Litz, & Hofmann, 2011; Propper et al., 2005), whilst men have been shown to disproportionately die by suicide (e.g. Uutela, 2010; Barr *et al.*, 2012; Reeves, McKee and Stuckler, 2014). Age has also been shown to have an association with mental distress, suggested to follow a U-Shape[2] over the lifecourse (e.g. Blanchflower & Oswald, 2008; Jorm et al., 2005; Puustinen et al. , 2011). Marriage has been shown to have a protective effect against mental illness (e.g. Law & Sbarra, 2009; Lindstrom & Rosvall, 2012; Wilson & Oswald, 2005). Ethnicity has been suggested to have a more complex relationship with mental health, with the effects differing depending on the neighbourhood status (Ross, Reynolds and Geis, 2000), but broadly the weight of evidence seems to suggest that non-whites experience worse mental health (e.g. Lang et al., 2011; Propper et al., 2005).

A number of socio-economic indicators have been shown to have an association with mental health. The true cause for this association is contested and is often suggested to be a manifestation of the relationship between perceived security, discrimination or mistrust and mental illness (e.g. (Jones-Webb and Snowden, 1993; Mays and Cochran, 2001; Whaley, 2001; Cairney and Krause, 2005; Corrigan *et al.*, 2012; Oguz, Merad and Snape, 2013). This is complicated further by suggested relationships between mental health and any individual socio-economic indicator tending to be hard to interpret in isolation due to the high correlations between different socio-economic indicators. Nevertheless, low socio-economic status (SES) has been consistently and systematically demonstrated to predict mood disorders (Lorant *et al.*, 2003), as well as psychoses (Argyle, 1994). Composite metrics of SES have been investigated and low-SES linked to depression (e.g. Propper et al., 2005; Ritsher, 2001). Lack of educational qualification has been shown to be associated with higher levels of mental distress (e.g. Grundy

[2] Specifics of the U-shaped curve are debated in the literature as possible artefacts of modelling processes incapable of separating age, period and cohort effects, see Chapter 2 or Bell, 2014 for a fuller explanation.

& Sloggett, 2003; Hu et al., 2007). Unemployment has also been shown repeatedly to be both a cause and effect of poor mental health (e.g. Ezzy, 1993; Taris, 2002; Flint *et al.*, 2013) A critical indicator of SES seems to be housing tenure, with non-home ownership associated with higher rates of mental illness (e.g. Levinson et al., 2010; Meltzer et al., 2012). Further complexity is added once causality is addressed, as there is no clear and unidirectional causal relationship. Research is inconclusive on whether low-SES causes depression or vice versa over the lifecourse - it is almost certainly a product of the two. As such, the association is clear but the reason for it is not fully understood (e.g. Muntaner *et al.*, 2004; Flint, Shelton, *et al.*, 2013).

1.5.2 The Presence of Geographical Gradients in Mental Health:

Moving beyond the strictly socio-demographic determinants of mental health there are also geographical relationships suggested in the literature that will be evaluated in this thesis.

Whilst there is a longstanding literature on the geographical gradient of physical health outcomes in the UK, mental health has yet to receive similar attention (e.g. Cruise & O'Reilly, 2015; Milne, 2014; Moller, Haigh, Harwood, Kinsella, & Pope, 2013). The North-South divide that is so often reported in UK physical health literature is under-evaluated in mental health literature (Duncan et al. 1995). Similarly to physical health, it has generally been found that there is worse mental health in Northern regions of England (e.g. Lewis & Booth, 1992), but this has been confined to objective measures. There is little on self-rated mental health differences between the North and South of the UK. Once subjective measures of wellbeing and mental health are considered the relationship becomes even less clear. Life satisfaction and wellbeing studies have found inverse relationships to those suggested in physical health literature, such as higher life satisfaction in Northern England and Scotland than London (Oguz, Merad and Snape, 2013). These studies seem to echo notions in positive psychology of subjective wellbeing's association with the social comparison model of mental health(Keyes, 2002; Slade, 2010; Ryff, 2013). The social comparison model posits that wellbeing commonly reflects an individual's *perceived* position in a perceived social, financial or health hierarchy, *not* their true position in these structures (Stutzer *et al.*, 2004; Dolan and Peasgood, 2006). This is not a new notion, for instance, relative earnings are commonly accepted as a more important predictor for mental health than absolute earnings (Stutzer *et al.*, 2004; Slade, 2010). That is to say, it is more desirable to earn £50k in an area where average income is £25k, than to earn £100k in an area where average income is £250k (Solnick and Hemenway, 1998). It is clear for this reason that modelling geographical differences in mental health presents greater methodological issues than modelling geographical differences in objective physical outcomes.

At a smaller scale, local environment has been found to have a significant effect on mental health (Evans, 2003; Galea et al., 2005). This has been suggested to be a product of several factors, including proximity to green space (Sugiyama *et al.*, 2008), quality of the built environment (Evans, 2003; Galea et al. 2005) and accessibility of local services or environment (ONS: Randall, 2012). Mair et al. (2008), conducted a review on 45 papers addressing neighbourhood effects on mental health and concluded that 37 of these studies did find significant effects of the neighbourhood on individual mental health but some clearly did not.

Fundamentally, it seems that where an individual lives matters for their mental health but it is not clear why. Associations between geography and mental health are well-established but whether this is simply an artefact of their composition or truly a contextual effect presents well-catalogued methodological challenges (Manley, Flowerdew and Steel, 2006; Mair, Diez Roux and Galea, 2008). The second half of this thesis will use a realistically complex modelling framework to address this to identify areas and places of elevated risk of mental distress. Furthermore, the thesis will try to see for whom geography matters most, as there is no reason to assume uniformity in area effects for different demographic groups and understanding this differential vulnerability would be beneficial for policymakers.

1.6 Thesis Outline

This thesis is divided into two distinct halves. The first concerns measurement and interrogates what is truly being captured by, and what can be learnt from, existing mental health outcomes. Much of previous work has adopted a relatively simple approach to mental health, treating it as a single, unidimensional construct. This is either captured by objective clinical outcomes, which are shown to under-represent mental illness considerably, or captured by inference from unidimensional summed scores of mental health screening questionnaires. These simplistic approaches presume that both the measurement and the assumed construct underpinning it are both reliable and reproducible. This thesis interrogates these methodological assumptions and examines multiple underlying dimensions within mental health measures. Furthermore, it evaluates the extent to which wellbeing and mental illness can be conceived of as simply differing ends of a single mental health spectrum. Given this more detailed approach to both measurement and modelling, even well-established research has to be re-evaluated.

The second half concerns the geographical and demographic patterning of responses to these outcomes in the UK in the years since the 2008 global recession. Using novel and advanced methodological techniques it explores which demographic groups and geographical areas are suffering disproportionate mental health burden. It then goes on to further examine the relationships, seeing if geography matters equally for all and if the worsening mental health suggested across the UK in recent years is affecting all individuals equally. These methodological contributions will help develop mechanisms to address these inequalities in mental health and gain insight into how to approach the complex topic of improving mental health in the UK.

This thesis aims to provide a comprehensive overview of the current state of mental health in the UK, both in terms of the current trends, and how policy can most effectively monitor and improve UK mental health. Using empirical evidence from a large-scale nationally representative survey, this thesis will use advanced modelling techniques to identify the determinants of mental health in the UK, as well as critically evaluating the current methods by which it is monitored and viewed.

1.7 Research Questions

In terms of the overall thesis several macro-scale guiding research questions present themselves.

1. What is mental health?
    a. What is it that is being truly captured when considering traditional mental health measures such as the GHQ-12?
    b. Does this interpretation fit with how it is commonly deployed in quantitative studies for policy development?
2. Is mental illness simply the inverse of mental wellbeing?
    a. How does the structure of the GHQ-12 relate to newly developed wellbeing measures such as the SWEMWBS?
    b. How are the two scales empirically related; and to what extent are the underpinning mechanisms driving each related?
    c. Given an increased understanding of complexity, can more dated measures such as the GHQ-12 be reinterpreted considering newer understandings of mental health?
3. Once these measures of illness and wellbeing have been appropriately interrogated, what are the key socio-demographic determinants of positive and negative mental health in the UK?
    a. Which demographic groups are most at risk of poor mental health and low wellbeing?

4. What effect does geography have on mental health?
    a. Is the effect of geography consistent for both positive and negative mental health?
    b. What spatial scale is most important for mental health intervention?
    c. For which demographic groups does geography matter most?
5. Incorporating our greater understanding of what is truly being captured, with greater methodological capacity for characterising demographic and geographical variation, is it possible to develop a more realistically complex understanding of the nation's mental health in the UK following the 2008 Global Financial Crisis?
    a. How temporally consistent is mental health? Whilst it is acknowledged that aggregate rates of mental illness in the UK are rising, are the individuals displaying depressive symptoms in a given year the same in the following year?
    b. Considering longitudinal data, which groups are suffering disproportionately in the years following the 2008 GFC?
    c. Are these groups consistent across the newly developed differing dimensions of mental health?
    d. Do all demographic groups follow the same trajectories over the period?
    e. If not, which groups are most at risk of poor mental health and low wellbeing?

1.8 Unique Contribution of this Thesis

This thesis looks to make several unique contributions to the current literature on UK mental health, both substantively and methodologically.

Substantively, the research looks to unpack traditional mental health measures, using newly developed methodologies, contributing to current debates in mental health understanding. It then incorporates these decomposed dimensions directly into longitudinal modelling as responses to try and disentangle the underlying processes determining mental health. Previously, studies have not tended to move past simple aggregate interpretation of mental health measures, such as questionnaires, taking the measures at face value, rather than interrogating their underlying characteristics.

Moving beyond specifics of responses, previous studies on mental health have tended to not have a comprehensive spatial structure, despite the acknowledged importance of local and neighbourhood context. This thesis will not only evaluate mental health with a thorough spatial structure, but also moves away from the strictly hierarchical conveniences of traditional multilevel analyses, into comparative modelling of locational and functional geography. This allows the research to move towards determining whether it is the specific place in which you live which is responsible for these contextual effects, or whether it is something inherent about the types of places, aside from strict geographical proximity. Finally, incorporating all these innovations, this thesis culminates in a comprehensive overview of UK mental health in the period 2009-2016. This is carried out using two extensively validated mental health measures, decomposed into their underlying dimensions via factor analysis. The deconstructed constructs of the illness measure are then taken as responses for the years 2009-2015, allowing the different elements of mental health to be modelled individually over the period directly following the global economic recession of 2007-8.

Methodologically, this thesis integrates several novel approaches to health modelling, which have not been previously employed in the field of mental health research. It also serves a guide for future testing and modelling of large scale questionnaire responses. The work in this thesis highlights the synergy of two cutting edge modelling techniques in Structural Equation Modelling and Multilevel Modelling and demonstrates how they facilitate the capturing of heterogeneity between measures within the individuals, whilst maintaining the capacity for generalisability at the aggregate level. Furthermore, this heterogeneity can be structured or decomposed with respect to a number of demographic and geographical indicators, allowing a detailed view of the mental health of the UK highlighting who and where is suffering the most.

1.9 Data

The scale of the investigation of this thesis is orders of magnitude larger than the sample sizes typically seen in mental health questionnaire literature. This is made possible by the inclusion of mental health questions in the UK Understanding Society Dataset (US). Formerly the British Household Panel Survey (BHPS), US incorporated and overhauled the BHPS, expanding its sample size from 8,000 to 50,000. Steps have been taken to ensure its continuing representative nature, so that findings from the data can be generalised to the population of the UK.

Initial cross-sectional analyses use the 2009 Wave, as at the time of writing it was the only released wave with data for both GHQ-12 and SWEMWBS[3]. Six waves of US data are available at the time of writing, constituting 240,000 responses to the GHQ-12 over the period 2009-2016 with each wave containing data collected over a two-year period. GHQ-12 responses were included as individual items, as well as a derived score, for all waves. SWEMWBS data for both total score and individual items is available on individuals for the waves 1 and 4. Furthermore, the same individuals who responded to the self-reported questionnaires will have the option of answering both responses, allowing greater inference

[3] Wave 4 also had both measures, but became available later

about the underlying mechanisms for each response. For instance, in 2009, 39700 individuals responded to the GHQ-12 and of those 37836 also responded to the SWEMWBS[4].

The data provides geographical information on each respondent, allowing the modelling of their household and local neighbourhood, with respondents selected via a clustered sample of UK postcode sectors[5]. Each of the selected postcode sectors had a stratified sample of households selected based on broad geographical regions, subsequently stratified by proportion of non-manual workers, then stratified by population density and finally ethnic minority density. A systematic random sample for address selection within these selected sectors gives us a final sample of 49,915 households, within 2640 Primary Sampling Units (PSUs) (McFall, 2011)[6].

To complement this unusually rich geographical data, information on area classifications is combined in analyses in Chapter 4 to investigate the impact of individual context both in terms of specific location as well as type of location (Vickers and Rees, 2006). US also collects data on over 1600 variables from each respondent, several of which are included in analyses as potential predictor variables. These include; sex, age, ethnicity, marital status, job classification, highest educational qualification and housing tenure. These are all cross referenced via individuals being given a cross-wave identifier, allowing them to be tracked through subsequent waves of data collection.

---

[4] For an idea of scale, the original validation of the WEMWBS response was carried out on a sample of 348 individuals.

[5] This is the case for England, Scotland and Wales. Northern Ireland addresses were selected separately using an un-clustered systematic random sample from the Land and Property Services Agency.

[6] Throughout the thesis this data is occasionally supplemented with other cross-referenced geographical data, for example from the population census, but these are appropriately cited when used.

1.10 Methodology:

To carry out these aims, methodological techniques are required that can answer the complex questions being asked in this thesis. It is dealing not simply with observed characteristics, but a combination of observed and latent characteristics operating simultaneously at several spatial scales. A very brief overview of these is given below, with more detailed discussion and context-specific justification provided in subsequent chapters.

1.10.1 Factor Analysis

Factor analysis and associated Structural Equation Modelling (SEM) approaches use multiple responses as indicators of the outcome of interest as an underlying trait. These approaches specifically model what is being captured via this underlying unmeasured or “latent” construct via the associated correlations in the questionnaire items (Asparouhov and Muthen, 2009). These latent constructs are referred to as “factors”.

For complex entities, such as mental health, it is not uncommon for these responses to have multiple underlying latent structures (e.g. Graetz, 1991; Kalliath, O’Driscoll, & Brough, 2004). These underlying constructs can also be statistically related to each other. In previous work the most common method of evaluating an underlying structure in a multi-item response has been Exploratory Factor Analysis (EFA) to suggest a structure followed by Confirmatory Factor Analysis (CFA) to corroborate it. However, it has previously been a requirement of CFAs that no single item can be considered a manifestation of more than one underlying factor (Jöreskog, 1967). This assumption of orthogonality in response items has been demonstrated recently to be unrealistic in social research and in fact can adversely affect analyses by artificially inflating factor correlations (Marsh *et al.*, 2014). This is critically important in understanding what is being captured by these mental health metrics as high factor correlations are often cited as reasons for assuming a simpler model (e.g. Gao et al., 2004; Gouveia et al., 2010; Padrón et al., 2012; Romppel et al., 2013). In 2009, Asparouhov and Muthen developed a method which allows the relaxation of this assumption which they termed Exploratory Structural Equation Modelling (ESEM). Crucially this allows us to move away from oversimplifications of mental constructs and start to engage with realistically complex understanding of the underpinning factors rather than impose too simple a solution.

There is a large body of research on the factor structure of the GHQ-12 and similar mental health responses, however most research tends to use the restrictive CFA specification. Using the newly innovated ESEM technique this thesis will offer both substantive and empirical improvements on previous work. Substantively, relaxing the assumptions of CFA allows us to appreciate more fully the inter-relatedness of mental health constructs and avoid oversimplifications. Empirically ESEM allows us to gain better fit and reproducibility in models as well as better representing the full range of dimensions captured by mental health measures without artificially inflating their factor correlations.

1.10.2 Multilevel Modelling

Also known as random-effects models, hierarchical linear models or mixed models (Goldstein, 2011, Longford, 1993, Raudenbush & Bryk, 2002), multilevel models are suggested to be the "gold standard" of contextual analysis (Chaix *et al.*, 2005). The multilevel approach, as adopted here, views individuals as nested within their household and neighbourhood areas and is designed specifically to handle data with a nested structure such as the US survey data. Due to the contextual understanding offered by the nested structure, the inter-dependence between individuals within the same household or same neighbourhood is taken into account in the modelling process. The approach simultaneously models the individuals' personal characteristics via a micro-equation whilst also allowing for higher level contextual effects such as neighbourhood characteristics in a macro-equation. These multiple levels are combined and estimate as a single overall model.

In 2008, Mair et al. conducted a review of 45 previous studies investigating the presence of a neighbourhood effect on mental health and concluded that 37 of these studies suggested the presence of one. As such and given that specific mental health processes have been demonstrated to be present at multiple geographical scales (e.g. (Leventhal and Brooks-Gunn, 2003; Galea *et al.*, 2005; Propper *et al.*, 2005) adopting an approach which can specifically model these seems imperative. Furthermore, in explicitly modelling these multiple scales simultaneously, these models avoid atomistic and ecological fallacies of inferring aggregate characteristics from the individual or vice versa (Diez-Roux, 1998).

This approach allows the modelling of differences in response at multiple spatial scales, which allows inference about which areas are associated with mental health at any of these scales. In doing so, it improves estimates of the neighbourhood or contextual effects and allows decomposition of variance into scale-specific values, thus allowing estimation of the relative importance of each level. Furthermore, it is readily extendable to include repeated measures, with the inclusion of an occasion level nested within the individual, which allows the investigation of variations in mental health in both space and time simultaneously (Jones, 1991). This in turn allows the modelling of both within-individual, between-observation personal change, and broader scale contextual change (Hoffman, 2015).

1.11 Thesis Structure:

As stated earlier, the main body of this thesis is split into two halves. Each of these halves contains two full research chapters, an outline of which is given below. Broadly the first half concerns itself with measurement, with the chapters using exploratory structural equation modelling to analyse what is captured by a traditional illness and wellbeing questionnaire, and the degree to which both the observed variation and underpinning constructs of each mental health metric are related. This investigates what can be known as mental health measurement and deals with the crucial question of the separation of processes of mental health into illness and wellness.

The second half concerns itself with the demographic and geographical patterning of mental health in the UK. Having understood more fully what is being captured by the mental health responses thanks to the first half, this goes on to characterise their distribution across the UK. It first looks at what can be known when considering the mental health metrics as they are intended to be used, as summed, unidimensional constructs. The final chapter then models responses to the differing components of the illness measure to see what can be known if the assumptions evaluated in the first half are integrated into the modelling framework introduced in the first section of the second half of the thesis.

Whilst this thesis is made up of thematically consistent chapters throughout, these are also designed to be read as standalone contributions to allow subsequent publication as journal articles. Consequently, there is some inevitable repetition at points of methodological comparisons and data sources. These proposed papers along with their distinct contributions are detailed below:

1.12 Mental Health Measurement

1.12.1 Chapter 2:

The General Health Questionnaire is globally one of the most widely used mental health metrics for population screening and quantitative research. It was initially developed in 1972 as a semi-diagnostic measure, intended to highlight individuals at risk of elevated psychological distress, particularly psychoses (Goldberg, 1972). Due to its enduring popularity and early innovation, there is a great wealth of information collected on the GHQ-12 (Goldberg *et al.*, 1997). The UK is no exception to this, with longitudinal datasets such as the British Household Panel Survey (BHPS) including the GHQ-12 in its data collection since its first wave in 1991. Despite this global and lasting usage, there is still much debate as to what the measure is actually capturing. In light of newer understandings of positive mental health, and given its persisting usage, researchers need to be sure that the GHQ-12 is still fully capturing valid dimensions of mental health. It would be simple to dismiss the GHQ-12 considering these newer understandings; however, given the wealth of data available for the GHQ-12, it is important to investigate the capacity for reinterpretation of what it is capturing, rather than disregard this wealth of information. This chapter aims to address whether there are methods by which the GHQ-12 can be reinterpreted, using newer understandings of mental health and new factor analytical methodologies to give a more rounded assessment of mental health.

Typically, the GHQ-12 is interpreted as an additive scale, with thresholds that are conceived of as denoting “elevated risk”. This chapter scrutinises and evaluates the underpinning assumptions behind this usage. Summing the questionnaire responses carries the inherent assumption that all the questions are capturing the same, singular, underlying dimension of mental health. Furthermore, analyses which stick to this rigid summed scoring make further assumptions of measurement error in the GHQ-12, assuming that individuals represent their mental health accurately across each item, rather than attempting to uncover consistencies and uniqueness within each of the constituent items.

This chapter uses methodological and inferential advances from recently developed Exploratory Structural Equation Modelling (ESEM) techniques (Asparouhov and B. O. Muthén, 2009; Marsh *et al.*, 2009), along with newly developed Bayesian implementation (Muth, 2010; Asparouhov, Muthén and Morin, 2015), and a sample size an order of magnitude larger than typical factor analytic studies (N=40,542). Utilising this improved methodological capacity, the underlying factor structure of the GHQ-12 is interrogated more comprehensively than ever before in a UK context using data from the 2009 wave of Understanding Society.

This is not the first time that this has been attempted, but existing debates on the topic are hardly conclusive. Typically, they have been carried out using Principal Components Analysis (PCA), which imposes rigid constraints on the possible factor structures found. Furthermore, the resultant factor structure has been constructed using a sample size an order of magnitude larger than that typically found in these studies (N=40,000). Finally, ESEM allows multiple numbers of underlying factors to be related to each item of a questionnaire, which has been previously impossible in traditional CFA. This has led to artificially inflated factor correlations, which have been used to justify the perceived uni-dimensionality of the GHQ-12 factor structure. By relaxing these stringent assumptions, a more realistic view is given of the complex and numerous constructs that underpin individual mental health. This chapter presents for the first time, a robust four-factor solution, made possible by methodological innovations in ESEM. This will allow researchers to characterise individuals more accurately, with multiple axes on which to be measured, reflecting the realistically complex nature of mental health measurement. The conceptual and empirical benefit of adopting this structure is then demonstrated as it is carried forward into subsequent chapters as a series of more representative and comprehensive dependent variables than a singular summed GHQ-12 score.

1.12.2 Chapter 3:

Continuing the full investigation of mental health measurement and evaluating conceptual inference, this chapter develops on the methods from Chapter 2 and applies them to a well-used wellbeing metric to evaluate similarities to the GHQ-12.

Whilst the GHQ-12 has been shown to be a robust measure of mental health, understanding of mental health has improved and broadened considerably since its introduction in 1972. Specifically, this has been characterised by the appearance and subsequent prevalence of the notion of "wellbeing", as a measure of positive mental health. This notion of positive mental health is one which is not necessarily even characterised as strictly related to negative mental health (Westerhof and Keyes, 2010; Keyes and Simoes, 2012). This carries important implications for research and policy as findings from illness literature are commonly translated into recommendations for the promotion of wellbeing. Against a backdrop of research increasingly moving towards an understanding of health beyond absence of illness, this chapter interrogates the GHQ-12 with respect to an alternative wellbeing measure, the SWEMWBS, to see if the latent constructs share any commonality. In short, this chapter investigates quantitatively and comprehensively whether mental wellbeing can be conceived of as simply the inverse of mental illness, or whether it should be regarded as a separate construct.

This chapter serves to contrast the GHQ-12 with a more recently introduced wellbeing measure (SWE) (Tennant *et al.*, 2007; Stewart-Brown *et al.*, 2009). Developed in 2007, the SWE was designed specifically to address shortcomings in the GHQ-12 and similar measures and was validated against it in its original development. It specifically aims to catch positive mental health elements via hedonic and eudemonic wellbeing. Whether these elements are truly captured in the underpinning processes of the SWEMWBS is investigated using the same ESEM developments in Chapter 2, generating a novel factor structure for the SWEMWBS. It is important to go beyond simply regressing the observed scores for the summed metrics, to interrogate the similarities between these underpinning 'processes' driving the scores of each metric. Thus, this new structure and interpretation is then evaluated in several conceptual steps against the structure of the GHQ-12 developed in Chapter 2.

This chapter uses the same data taken from Wave 1 of Understanding Society. The majority of individuals who responded to the GHQ-12 questionnaire in 2009 (39700), also completed the SWE questions (37836). By comparing the same individuals across both measures, we can ascertain to what degree the underlying factors of the GHQ-12 are determining the responses for the SWE. Using the newly innovated factor structure for the GHQ-12 developed in Chapter 2, the similarity between the GHQ and the SWE is investigated. Regressing the individual factor scores for each of 4 factors on the individual responses for the SWE outcome allows us to see relations between each item of the SWE and the four underpinning factors of the GHQ-12. After interrogating the relationship between each of the individual responses to each questionnaire on the underpinning constructs of the other, the final step is to regress the underpinning constructs of either metric on each other. This provides an empirical evaluation of the degree of similarity between wellbeing and mental illness as captured by these two questionnaires.

This chapter contributes to a growing body of literature on the state of mental wellbeing research and contributes to the debate on the inter-connectedness of wellbeing and mental illness. Whilst again being carried out using a relatively very large sample, it also carries the methodological advances in Chapter 2 forward, whilst now contrasting them with the SWE. To the authors' knowledge this has never been achieved before. It is innovative in the scale of the analysis and in the methods used, and also in not oversimplifying the underlying factor structure of complex mental health issues. Traditionally, evaluation of similarity tends to remain solely in the manifest domain in large-scale quantitative mental health research, where this chapter goes interrogates the relatedness of the underpinning processes, having accounted for item-specific measurement error.

1.13 Mental Health distribution

Having gained a more comprehensive understanding of what is being captured by both the GHQ-12 and the SWEMWBS, the thesis now moves into investigating their distribution in the UK. It is well accepted that there is inequality in the experience of mental health. This inequality is patterned with respect to both demographic characteristics of individuals, and the geographical context in which those individuals are situated. Moreover, these two scales of patterning are related, with demographic characteristics being related to geography, thus the next two chapters must specify a methodology which can model both demographic characteristics and geographical context simultaneously to fully understand their context-specific effects.

1.13.1 Chapter 4:

The relationship between individual social status and physical health outcomes has long been acknowledged in epidemiological literature. Unlike its physical counterpart, however, mental health has not enjoyed the same consensus as discussed earlier. This chapter models mental health responses from both GHQ-12 and SWE with respect to a number of demographic and structural predictors in order to ascertain which groups experienced the best and worst mental health in the UK in 2009.

It is clear that the experience of mental health is not uniform across the UK (Weich, 2003; Propper *et al.*, 2005; Weich, Twigg and Lewis, 2006; Davies, 2013). The patterning of this variability geographically and demographically is investigated in this chapter. This understanding of the importance of geographical context can be operationalised most effectively using a bivariate (multiple outcome) multilevel approach, partitioning variance into the spatial scale at which it is most consistently expressed. Existing studies of the contextual effects for mental health are often contradictory, although overall it seems clear that there are more cases in which a contextual effect is found than not (Ross, Reynolds and Geis, 2000; Mair, Diez Roux and Galea, 2008; Fone *et al.*, 2013).

Of particular interest when investigating the importance of geography is characterising the degree to which areas are important because they are a specific location, or because they carry similar characteristics of deprivation. To evaluate this relative importance of locational or functional geography this chapter moves beyond the strictly hierarchical conceptions of geography that are traditionally analysed in multilevel literature. It specifies models which simultaneously consider geography as a locational process and a functional process. Each household is modelled as simultaneously belonging to both a specific geographical area, and a type of area, using a cross-classified structure. Furthermore, as this is a bivariate model, it allows the estimation of the similarity of the GHQ-12 and SWEMWBS at each structural scale.

This chapter further aims to evaluate what it is that makes individuals differentially vulnerable to psychological distress. To ascertain true estimates of geographical variability it is vital to ensure that estimated geographical differences are not simply artefacts of areas containing greater numbers of specific demographics. In order to characterise this, demographic predictors are included in the models, with geographical differences estimates having taken account of individual characteristics. Furthermore, specifying models with demographic predictor variables facilitates the analysis of the predictive capacity of several different social characteristics related to both the individual and household. A wide variety of individual characteristics are investigated, all of which have been previously found to have a relationship with mental health.

This chapter contributes a different but far more comprehensive approach to conceptions of mental health, moving away from strictly locational geography, investigating area effects net-of compositional effects, and simultaneously modelling both responses to investigate their similarity across each structural level. This allows us to ask, at what structural level a positive outcome for wellbeing most implies a positive outcome for mental illness, which offers valuable insight into the most effective levels for policy intervention. This chapter concludes with an overview of which specific geographical locations, which types of locations (characterised by area classifications), and which demographic groups suffer the greatest burden of mental illness. Furthermore, it characterises geographical variability with respect to age, to ask at what age does geography matter most, as a demonstration of the necessity for nuanced interpretation of these findings.

1.13.2 Chapter 5:

Chapter 5 is the most ambitious chapter in the thesis. It combines the methodological contributions of all the previous chapters, into a comprehensive overview of the mental health of the UK in the years since the economic recession. It incorporates longitudinal measurements of the GHQ-12 over 6 waves of Understanding Society, consisting of over 240,000 observations. It is principally concerned with chronicity – that is the degree to which mental health as captured by the metrics is temporally consistent within an individual. More simply, this chapter investigates whether the ill stay ill and the well stay well. This longitudinal data, coupled with a multilevel approach to longitudinal data modelling allows the investigation of the relative importance of the individual, as opposed to the occasion at which the observation is made. That is – whether mental health should be thought of in terms of “ill individuals” or “distressing time-periods”. This is of critical policy relevance, as if individuals tend to become chronically ill then it could be argued that mental health policy should favour individualised treatment similar to any other long-term limiting illness. If the inverse were true, and the occasion of measurement were the most important factor, then it could be argued that policy should be chiefly concerned with mitigating environmental stressors.

The principal aim of investigated chronicity is supplemented with aims provided and informed by the methodological contributions of the previous chapters. The four-factor structure of the GHQ-12 identified in Chapter 2 is used to transform the unidimensional interpretation of the GHQ into four separate response variables, allowing for the capacity to determine separate patterns for each dimension of the GHQ-12. Furthermore, it is of interest to know to what degree these dimensions are correlated at each spatial scale, therefore these dimensions are modelled simultaneously in a multivariate specification.

This chapter builds on the framework of Chapter 4, using repeated measures within individuals as a new structural level. This involves the conceptual nesting of occasions within individuals. As outlined above, this allows the characterisation of the relative importance of the occasion and the individual, but across each of the four dimensions of the GHQ-12 outlined in Chapter 2.

The multilevel approach readily extends itself to the investigation of geographical and demographic variation in a manner like Chapter 4. However, where Chapter 4 investigated complex geographical structuring of variance, this chapter considers the relative importance of individual and occasion-level variance as the prime focus. The demographic investigation allows the evaluation of the findings of Chapter 4, in a longitudinal perspective – analysing whether the cross-sectional patterns hold longitudinally over the period.

There is an advocation of nuance throughout this thesis, which includes the consistent relaxation of assumptions of uniformity in mental health. This is due to the demonstrated complexity within mental health as a topic, and the acknowledged degree of complexity within the experience of individuals. It seems fitting then, to not assume uniformity in the mental health trajectories of different demographic groups over the 6-wave period. To this end, interactions are specified for each demographic predictor and the wave of measurement, allowing the separate estimation of the longitudinal trends of each demographic subgroup. This allows the identification of those groups who are experiencing a disproportionate burden of mental distress over the period, with obvious policy relevance.

This chapter contributes to the ongoing body of literature about the effects of recessions and subsequent austerity measures on mental health. This constitutes arguably the most comprehensive analytical method used to understand UK mental health to date. It characterises change between individuals across four-responses, each allowed to vary separately, but retain covariance in order to understand the similarity in process. It then structures this variation by individual, household and observation levels, allowing evaluation of the degree of chronicity within mental health experience. These variations are then in turn characterised with respect to key demographic characteristics, which are subsequently individually interacted with time to relax the assumption of uniformity of trajectory across groups. Overall this comprehensive overview of mental health development in the UK will allow the addressing of questions of causality, chronicity and distribution of underpinning mental health processes as never before.

1.13.3 Chapter 6:

The thesis concludes with a chapter outlining the main findings of the previous five chapters which constitute the main body of work. Chapter 6 details some of the limitations of the study and poses important questions which need to be addressed in future work in light of the thesis findings.

1.14 Policy Relevance and Impact:

This research will be of potentially great interest to a number of parties and preliminary results have already elicited interest from Mind, the largest mental health charity in England and Wales, who requested a presentation for their policy-influencing team on the early findings of the work. Developing the most accurate method of understanding mental health is something that can be carried out worldwide, specifically the combination of factor analysis and multilevel modelling in a longitudinal framework. The methodological processes outlined in this thesis are readily extendable to other mental health measures by design, meaning that the process could equally be applied in other settings to other commonly used mental health metrics. The response measures which are evaluated in this thesis are very widely used in UK based mental health literature. It has been observed that more dated mental health measures have been suggested to fail to capture newer “wellbeing” elements of mental health, however instead of discarding the wealth of information that has been collected on them since their conception, this thesis offers valuable insight into how this extensive existing data can be re-interpreted, re-interrogated and re-evaluated in light of new understanding of mental health. Additionally, the greater understanding afforded by the multilevel aspects of the thesis will help indicate at what spatial scale this increased understanding should be implemented into policy. Determining this is valuable for policymakers interested in improving inequalities in mental health. The same motive drives the interest in demographic predictors of mental health. There are clear and demonstrable motives for policy in developing a clear and comprehensive understanding of those groups within the UK who are suffering a disproportionate burden of mental distress.

This research then goes further, highlighting specific geographical areas which may have higher risk of mental illness and/or psychological distress. Providing detailed trajectories of the different aspects of mental health in the period following the economic recession has important policy implications. Seeing which dimensions of mental health have served to drive the mental health response of the UK since the recession and for whom they are most important will allow future policymakers to target specific groups for intervention post-crises, as well as provide insight into the short- to mid-term reactions to these shocks.

The investigation into the degree of chronicity within mental health is also a novel and important avenue for quantitative survey research. An understanding of the degree to which mental distress is temporally consistent will shed light on the most effective treatment and intervention options. If it is demonstrated that individuals are relatively consistent through time, experiencing enduring high levels of mental distress, then treatment should perhaps follow a model similar to that which is typically deployed for limiting long-term illness. However, if it is shown that the occasion is more important than the individual, that would suggest that individuals tend to fluctuate in and out of depressive states, which could be used to motivate greater investment in mitigating environmental ‘distressors’. Whilst it is almost certain that there will be an element of both, the relative importance of each is an avenue that is relatively unexplored in longitudinal analyses which typically just identify greater raw numbers of distressed individuals within each subsequent year, without consideration of whether those individuals are consistent between years.

Fundamentally this will provide valuable information on not just which geographical areas and demographic groups could be best served by preventative measures, but also provide insight into how these processes actually manifest. Are the processes driving mental illness the same as those driving mental wellness? Should findings from traditional mental illness research be translated into recommendations for improving positive mental wellbeing? To what degree should individuals suffering from common mood disorders be thought of as ill individuals, or victims of distressing circumstance? These questions are of critical importance in the ongoing debate about mental health and this thesis aims to provide insight and answers to those questions and inform how best to address the current inequalities in the mental health of the UK.

2 DEFINING THE STRUCTURE OF THE 12-ITEM GENERAL HEALTH QUESTIONNAIRE USING EXPLORATORY STRUCTURAL EQUATION MODELLING

2.1 Background:

2.1.1 Measuring mental health

The 12-Item General Health Questionnaire (GHQ-12) is one of a suite of instruments used to assess psychiatric morbidity both in community and clinical settings (Goldberg & Gater, 1997). The capability of the 12-item GHQ to consistently reproduce the results from its longer counterparts, coupled with its relative brevity for use in the clinical setting, makes it appealing as a measure of mental health to include in large scale surveys (Goldberg & Gater, 1997). Despite its initial focus on diagnostic purposes for specifically at-risk individuals, the GHQ-12 has since been translated and validated across multiple languages and countries as a screening tool for depression and depressive symptoms across populations (Hankins, 2008b; Smith et al., 2010). Its subsequent inclusion as a mental health measure in a number of large scale population surveys has consolidated its status as the most widely used screening instrument for common mental disorders (Werneke *et al.*, 2000). The UK is no exception to this. With the inclusion of the GHQ-12 in the British Household Panel Survey (BHPS) and later Understanding Society (US) as well as in numerous other surveys such as the Health Survey for England (HSE), the GHQ-12 has become a canonical mental health measure for UK-based population studies (e.g. Propper et al. 2005; Weich 2003; Pevalin 2000).

As outlined in Chapter 1, mental health, particularly that characterised by mood disorders, is notoriously hard to measure. This is due to its lack of easily quantifiable thresholds and tangible physical symptoms, resulting in severe under-diagnosis (Verheij, 1996). It has been suggested that up to 42-46% of depressed individuals do not ever receive recognition as psychiatric cases, and only 15-26% of depressed patients actually receive diagnoses of depression (Lecrubier, 2007). This under-diagnosis is again linked to severity, due to measurement difficulty, with the milder cases of mental illness being the most often under-reported (Garrard *et al.*, 1998). However, the development of mental health scales, such as the GHQ-12 (Goldberg, 1972), has helped research move away from more easily quantifiable psychoses which have more commonly been the focus of mental health research (e.g. Giggs 1973; Silver et al. 2002). The GHQ was developed as a diagnostic tool for mental illness with dual focuses: (a) on individuals' inability to carry out normal function, and (b) on the emergence of distressing phenomena (Goldberg and Williams, 1988). The shortened 12-item version has been shown to produce "remarkably robust" results in comparison with initial longer versions (Goldberg *et al.*, 1997). It has been extensively applied and validated across multiple populations for the measurement of general psychological health (Pevalin, 2000; Creed and Evans, 2002) and has become the metric of choice for measuring mental health.

Due to its widespread uptake, the GHQ-12 has also become the standard outcome for quantitative analyses of mental health. In the years since its development, however, attitudes towards mental health have matured, resulting in the progression of GHQ-12 usage from individual-level diagnostic function to population-level screening. Additionally, increased awareness and promotion of mental health has seen depression and mood disorders become outcomes of interest in research, in turn increasing need for diagnostic tools beyond clinical diagnoses. This can be viewed more broadly as part of a larger shift, often cited as a move from an "absence of illness" perspective, which focused on health improvement as achievable via the reduction of clinical illness diagnoses, to a more holistic promotion of positive mental health, which considers all individuals to lie on a mental health continuum with room for improvement (Davies, 2013; World Health Organization, 2013). Despite this broader appreciation of complexity, interpretation of GHQ-12 results and its inclusion in analyses have remained relatively basic, rarely moving past inference to a generic “mental health” mechanism using simplistic procedures of summed scores across items. Whilst the instrument was originally intended to be unidimensional, several factor analytical studies have suggested multiple two- and three-factor structures underpinning the response. Given the wealth of data available for the GHQ-12, re-evaluating what it is that is being captured by the items using improvements in contemporary methodology is crucial for furthering the understanding of measures of mental health offered by population studies.

This chapter initially presents an overview of the debates surrounding the traditionally deployed GHQ-12 with a summary of the implied numerical assumptions associated with its use. It then goes on to give a brief overview of the dominant factor structures previously identified in the literature, highlighting those which are subsequently compared against the structure identified here. This chapter then makes explicit suggestions for GHQ-12 interpretation, presenting a new structure for interpreting the GHQ-12 using recently developed Exploratory Structural Equation Modelling (ESEM) which allows for the relaxation of the more unrealistic numerical assumptions of previous research. Finally, this new structure is evaluated both empirically and theoretically against the most widely adopted interpretations of the GHQ-12 using a large scale, representative sample (N=40452) from the UK. The outcome is a set of validated constructs of differentiated mental health underpinning the GHQ-12 that can be used in future studies. Instruction on the construction of these dimensions is provided for future researchers, and the use of this methodology is encouraged, particularly for monitoring, analysing change, and answering questions about determinants of mental health responses.

2.1.2 From Questions to Concepts: What are we capturing?

The GHQ-12, as the name suggests, consists of 12 questions which are given in Table 2.1. Respondents are asked to give an ordinal Likert response considering their recent life, typically indicated as "the past few weeks" following good practice guidelines outlined in Cox et al. (1993). Scoring of responses differs based on the positive or negative nature of the item, but higher scores always indicate higher psychological distress. The responses to the 12 items are then typically summed to give a score between 0 and 36, with higher scores indicating greater distress as shown in the table.

GHQ-12 Item		***Scoring***		
"Have you recently?"	**0**	**1**	**2**	**3**
1. Been able to Concentrate on what you're doing?	Better than usual	Same as usual	Less than usual	Much less than usual
2. Lost much sleep over worry?	Not at all	No more than usual	Rather more than usual	Much more than usual
3. Felt you were playing a useful part in things?	More than usual	Same as usual	Less so than usual	Much less than usual
4. Felt capable of making decisions?	More so than usual	Same as usual	Less so than usual	Much less than usual
5. Felt constantly under strain?	Not at all	No more than usual	Rather more than usual	Much more than usual
6. Felt you couldn't overcome your difficulties?	Not at all	No more than usual	Rather more than usual	Much more than usual
7. Been able to enjoy normal day-to-day activities?	More than usual	Same as usual	Less so than usual	Much less than usual
8. Been able to face up to your problems?	More so than usual	Same as usual	Less so than usual	Much less than usual
9. Been feeling unhappy and depressed?	Not at all	No more than usual	Rather more than usual	Much more than usual
10. Been losing confidence in yourself?	Not at all	No more than usual	Rather more than usual	Much more than usual
11. Been thinking of yourself as a worthless person?	Not at all	No more than usual	Rather more than usual	Much more than usual
12. Been feeling reasonably happy, all things considered?	More than usual	Same as usual	Less so than usual	Much less than usual

Table 2.1: List of GHQ-12 questionnaire items with scoring for responses (Goldberg & Gater, 1997). Italicised questions indicate negative phrasing with responses altered to score inversely such that higher scores always indicate greater distress. Scores are typically summed across the 12 items to give a combined response ranging from 0-36.

Despite the widespread uptake of the GHQ-12 there is still contested ground as to what the measure is actually capturing (Werneke *et al.*, 2000; Ye, 2009). Broadly, arguments about the appropriateness of the measure can be divided into two distinct but inter-related issues, the first of which is scoring, that is the numerical coding of response categories. The second of these arguments regards dimensionality, the extent to which there is a single underlying "mental health" construct underpinning the responses or that multiple dimensions are required to capture the reality of mental health experience. Both aspects, as will become clear, are influenced by the methodological choices which determine what is and what can be found.

2.1.3 Response Scoring

Two main scoring structures exist for the GHQ-12, binary (0-0-1-1) scoring, and the unadjusted Likert (0-1-2-3) scoring displayed in Table 1. For each of the 12 items, individuals select from four ordinal responses for each item, these answers are inverted for negatively phrased items such that higher scores indicate greater mental distress. The originally recommended 0-0-1-1 binary scoring has been adopted by the majority of studies (Goldberg & Gater, 1997). This initial recommendation was primarily to deal with the problematic assumption of additivity; that the scores can validly be summed. Goldberg (1972) was concerned with the issue of "conceptual distance" between the response scores and this conceptual distance being different between different response categories within the same item. Their proposed solution truncates the information, if the categories of response are conceptualised as differences in likelihood of responding positively for each question, the binary scoring treats the only important change as that from "better than usual" to "worse than usual". Instead, the Likert scoring treats the differences between responses as informative regardless of the conceptual distance between the scores. The initial binary approach is predicated on unidimensionality, assuming that all 12 questions are dictated by a *single* underpinning dimension, with this conceptual distance having to be uniform across this single underlying construct. Recently, mental health research has started to move beyond assumptions of a single underlying dimension towards an understanding of mental health as underlain by more than one dimension. As this shift is realised, there is less of a need to impose such strict structure on GHQ-12 scoring in order to legitimise the additivity assumption, as each item is no longer necessarily measured along the same single metric. Hence alongside the development of multidimensional outcomes from the GHQ there has been increased adoption of Likert scoring (Andrich and Schoubroeck, 1989).

Options beyond binary and Likert scoring have been proposed, but not widely adopted in the literature. A cursory glance at the descriptive statistics in Figure 2.1, showing GHQ-12 responses for Wave 1 of Understanding Society (2009) illustrates the differences in response patterning between the positive and negative items. The positive items have a much stronger "floor" effect – that is the majority of respondents tend to state "Same as usual" to indicate lack of problem. This differs in the negative questions as there is some ambiguity in the response categories as to what the "normal state" response is meaning. Consequently, the modal response category for these items is far less clearly defined, with the responses instead spread between "Not at all" and "No worse than usual". This has resulted in suggestions of differentiated scoring methods for positive and negatively phrased items to try and reduce over-interpretation of these differences. Most notably an adapted binary scoring (0-1-1-1) was proposed by Goodchild and Duncan-Jones (1985) for negative items, claiming to more realistically capture individuals with existing psychiatric morbidity.

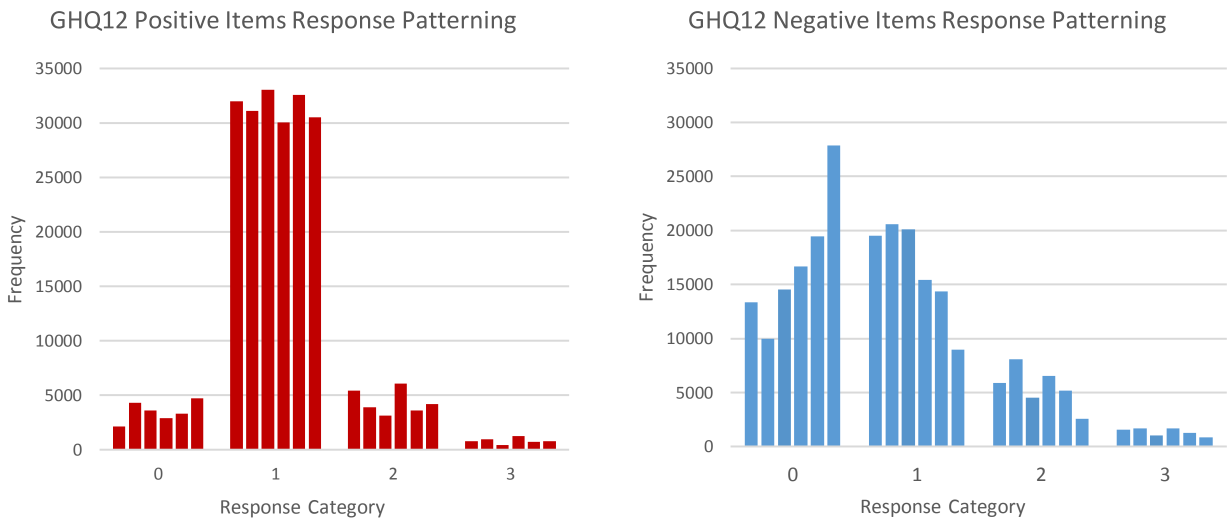


Figure 2.1: Graph illustrating the differential frequency of response patterning of positively (ACDGHL) and negatively (BEFIJK) phrased items in the GHQ-12 from 2009 Wave of Understanding Society.

Despite this, strong arguments have been made for the ordered, categorical, 0-1-2-3 Likert scoring, as this does not compress the information given by the participant, and has indeed been argued to be superior in modelling contexts as its contains more information (Winefield *et al.*, 1989; Campbell, Walker and Farrell, 2003). Moreover, Likert scoring has been argued to allow greater discrimination between different dimensional structures (Campbell & Knowles, 2007; Smith et al., 2013). Despite acknowledged differences in responses on positive and negative items, the "corrected scoring" proposed by Goodchild and Duncan-Jones (1985) solely deals with binary categorisation, and thus still compresses information. It is for these reasons that the full information Likert scoring is used here to retain full information and offer greatest power in evaluating alternative underlying structures.

2.1.4 The Persistent Myth of Unidimensionality

The treatment of the GHQ-12 as a composite measure, summing up responses across items to give a total score from 0-36, carries the implicit assumption of *unidimensionality* in which all items are treated uniformly. This is especially pertinent for literature that recommends or posits "thresholds" for the GHQ-12 above which an individual is considered a risk category (e.g. Baksheev et al., 2011; Goldberg et al., 1998; Tait et al., 2003), as it is critical that the unit increase that pushes an individual over a threshold implies the same change in mental status between different questionnaire items.

The use of summed scores and thresholds also implicitly carries the assumption of *additivity* (Marsh and Bailey, 1991). Inference from summed scores not only requires consistency *between* items, but consistency *within* items, such that a unit increase implies the same change in the mental health of an individual wherever it occurs across different items (Brodersen *et al.*, 2007). If these assumptions are not met, then it cannot be assumed that the measure is capturing equivalent differentials across individuals.

Despite the obvious importance of these assumptions for informing public health conclusions, they are among the most problematic and contested assumptions in the literature (Werneke et al. 2000, Picardi et al. 2001, Hankins, 2008, Smith et al. 2010). There is considerable debate as to the 'true' underlying structure of the GHQ-12. Indeed many structures progressing beyond unidimensionality have been suggested to give a more comprehensive understanding of the mental health processes underpinning differences between scores (e.g. Andrich & Schoubroeck, 1989; Campbell et al., 2003; Kalliath et al., 2004; Penninkilampi-Kerola et al., 2006; Smith et al., 2010).

The unidimensional approach of considering mental health as entirely captured by a 0-36 numerical scale is often argued to be reductive. More complex methodological understandings are commonly recommended for interpretation of mental health responses (e.g. Martin & Newell, 2005, Hu et al. 2007, Aguado et al. 2012). To try and progress understanding beyond a simple linear metric, factor analysis is often used to help identify whether what is captured is a result of multiple, more complex, dimensions underlying the observed questionnaire responses. It does so by assessing whether responses to the constituent items tend to correlate with one another. It then relates these items to broader underpinning mechanisms, assumed to be dictating these correlated item groups, which are usually modelled as one or more latent, continuous variables – or factors – which are then the true object of interest.

2.1.5 Moving Beyond Unidimensionality - Exploratory Factor Analysis

Once the conceptualisation of mental health moves beyond unidimensionality it is important to appreciate at the outset that different approaches to factor analysis have the potential to produce radically differing results. Typically, the process of factor analysis is carried out in two steps, firstly extracting and rotating the structure, and secondly evaluating it against other potential solutions. The extraction and rotation steps are termed "Exploratory Factor Analysis" (EFA). Extraction methods are discussed here, with a fuller discussion of rotation given in the Methodology section. Extraction can be carried out using two approaches, the common factor approach or the principal components approach.

The common factor approach specifies that each and every constituent item loads on each and every one of a pre-specified number of latent variables and seeks to explain the correlations between the observed item responses using these factors. This assumes that there is some underlying common factor on which all items load, but crucially posits that there are deviations from this corresponding to each constituent item. More simply, each item is allowed to have a "unique" variance in conjunction with the shared variance that is explained via the underlying factor (Schmitt, 2011). Results are typically reported as standardised regression coefficients for each item on each latent variable, commonly referred to as a "loading". This represents the best fit to the data for the specified latent factors. However, traditionally more structure has had to be imposed in the form of constraints to enable testing against other structures and simplify construct interpretation (Sass and Schmitt, 2010).

The second approach to extraction uses Principal Components Analysis (PCA) to extract the initial structure. This approach seeks to best summarise the data in the fewest total component variables, maximising explained variance in each of the response items. This more "concise" summary entails the assumption that the dimensions (*components)* themselves must be orthogonal, constraining them to be uncorrelated, meaning that no variance is shared between dimensions. Due to the lack of distinguishing between "shared" and "unique" variance, this implicitly presumes measurement without error. More simply, PCA summarises the entirety of the data, regardless of whether it seems atypical or erroneous (Conway & Huffcutt, 2003).

The Principal Components approach has been critiqued at length since its initial application in social research due to the problematic nature of these assumptions, inflated estimates of explained variance, and the fundamentally different underlying goals of PCA and EFA (Thomson, 1939, Ford et al. 1986, Conway & Huffcutt, 2003, Gorsuch, 1997, Widaman, 2007, Schmitt, 2011). This last point is the most critical and once understood explains why these two methods can have such differing results despite having seemingly similar aims. Fundamentally the common factor approach, which underpins factor analysis, seeks to explain off-diagonal covariances (which can be standardised to covariances) between the observed or manifest variables (here the constituent questionnaire items). It does so by positing the presence of underpinning latent variables that represent these covariances (Thomson, 1939). The diagonal elements of this covariance matrix, the raw variances, are of little importance when the aim is to account for the relationships between the off-diagonal elements. Conversely, the goal of PCA is to account for as much variance as possible in the raw scores of each element in the fewest number of components, often for the purposes of data compression (Smith, 2002). Thus, all the variance in the raw outcomes is retained to account for maximal amount of variance across the observed variables. In PCA explaining the inter-relatedness of the responses, via the off-diagonal covariances, is of lesser interest as long as maximal variance is explained (Widaman, 2007).

There has been lengthy debate about the relative merits of these approaches in social research. Initially PCA was favoured as a quicker, easier alternative to the common factor approach which required greater computational power, however this has obviously become considerably less of an issue in the years since the debate began (Gorsuch, 1990, Conway & Huffcutt, 2003). The common factor approach is more appropriate for social outcomes, as these are even more likely to be measured with error than the engineering outcomes for which PCA was initially developed (Costello & Osborne, 2005). However, whilst the weight of evidence seems to considerably favour the "common factor" approach, PCA still sees relatively frequent use in initial factor extraction, be it due to methodological canons, or lack of awareness of shortcomings, with considerable implications for results (Fabrigar et al. 1999, Schmitt, 2011). Regardless of which of these two approaches is chosen by the researcher, the next stage remains to evaluate the proposed solution.

2.1.6 Evaluating Solutions – Confirmatory Factor analysis

The second stage in Factor Analysis is to test the current solution against other potential factor structures. This has traditionally required refinement of the extracted solution using a method outlined by Joreskog (1969) termed "Confirmatory Factor Analysis" (CFA). This seminal work sought to specify a generalisable method for estimating factor structures using Maximum Likelihood Estimation. However, in order to evaluate the proposed solution, it must be reduced to a format called "simple structure" where all but the largest coefficient for each item is eliminated from the analysis (McDonald, 1987). This results in each item loading on only one of the underlying constructs, producing a modelled series of latent factors each conceived of as singularly determining a series of questionnaire items. Using this simple structure solution, several fit indices can be produced which can be used to compare the relative merit of different proposed factor structures. CFA has remained the dominant method for evaluating existing factor structures and work on the GHQ-12 is no exception.

Despite being termed "Confirmatory" practice, the use of CFA methods often seeks to achieve the model with the best fit statistics for the data, a procedure which tends to be iterative and conceptually exploratory. The strict requirement of zero cross-loadings in the simple structure specification has been argued to lead to a reliance on overfitting (via extensive model modification) when seeking the best-fitting model (Asparouhov & Muthen, 2009). This is evidenced by extensive use of modification indices, which involve the running of parallel specifications with slight adjustments and then reporting which of these gives the superior fit statistics (Schmitt, 2011). This approach has been criticised for producing inconsistent results in repeated samples and moreover seems thematically at odds with the purposes of CFA, which fundamentally aims to test a theory-driven a-priori structure (MacCallum et al. 1992). In these cases, where the extensive re-specification has essentially used CFA for exploratory purposes, the original EFA is often a more appropriate solution than the extensively modified solution (Browne, 2001, Marsh et al. 2014). Furthermore, methodological justification in many factor analysis studies has been suggested to be lacking, with important choices on extraction, rotation and simplification largely being based on tradition or previous work, regardless of its statistical properties (Fabrigar et al., 1999).

Beyond substantive issues with factor analysis, there are also subject-specific issues. The imposition of simple structure is particularly problematic in complex social outcomes such as psychological constructs, where multiple and inter-related underpinning processes are likely to give rise to any specific outcome. As Marsh (2014) puts it, "nonzero cross-loadings are inherent in psychological research"., but the exclusion of multiple loadings for constituent items has led to oversimplified latent factors (Marsh et al. 2014). Until relatively recently it was not technically possible to evaluate the relative merit of different proposed models when a constituent item was related to more than one underpinning dimension. Given that refinement to simple structure was necessary at the outset, its substantive implications were rarely scrutinised, however with the innovation of ESEM this is no longer the case (Asparouhov & Muthen, 2009).

Most critical when considering these proposed structures, however, is that the imposition of simple structure distorts the results of the model. Simple structure is an imposition of orthogonality between factor determinants regardless of whether this is what the initial EFA shows. Correlations between the factors themselves, which were initially extracted as loading on every single item, are artificially inflated as they are merely simplifications of the initial constructs which were specified *with* cross loadings. A number of ESEM simulation studies with more relaxed constraints on factor specification have demonstrated this, showing that traditional CFA methods erroneously inflate factor correlations purely as an artefact of the highly constrained and unrealistic model specification (Asparouhov & Muthén, 2009; Marsh et al., 2014).

This is especially pertinent with respect to the GHQ-12, as high factor correlations between GHQ-12 factors proposed in the literature are regularly cited as arguments for simplifying to unidimensional interpretation (e.g. Fernandes & Vasconcelos-Raposo, 2013; Gao et al., 2004;

Gouveia et al., 2010; Padrón et al., 2012; Romppel et al., 2013). This is not necessarily unreasonable; the reported factor correlations are indeed often very high. It is not uncommon for reported correlations in the literature between GHQ-12 factors to be greater than 0.9 (e.g. Cheung, 2002, French & Tait, 2004, Gao et al. 2004, Campbell & Knowles, 2007) or even 0.95 (e.g. Sweeting et al. 2009, Martin & Newell, 2005, Wang & Lin, 2011). Indeed, an analysis of eight existing GHQ-12 factor structures, including some derived by PCA, by Aguado et al. (2012) reports at least one factor correlation over 0.90 for all but one evaluated solution. One of the key benefits of the ESEM approach adopted here is that these stringent assumptions inherent in traditional factor analysis can be relaxed. By allowing for the specification of non-zero cross loadings, the researcher is in fact allowing far more realistic factor estimation which does not artificially inflate correlations.

2.1.7 Previous Factor Structures from the Literature

Despite a great deal of research into the factorial structure of the GHQ-12 questionnaire, there is no real consensus as to which factor structure most fully captures the response patterns. Indeed, Multiple possible structures have been suggested. Simple unidimensional structures have been posited and used by many researchers, commonly with an adjustment for positive and negative phrasing (Andrich and Schoubroeck, 1989; Winefield *et al.*, 1989; French and Tait, 2004; Hankins, 2008a; Ye, 2009; Aguado *et al.*, 2012). The adjustment for phrasing aims to explain the repeated identification of two factors categorised solely by positive and negative items, and concludes that this patterning of responses is simply a product of wording. It is our view that this is an overly simplistic, and that this would be a reasonable conclusion only if the positive or negative specification of the questions were chosen at random and consequently the wording was unrelated to what an item is measuring. Whilst this may be an issue in the grouping of these items by design, it is still unlikely that mental health is categorised as neatly as this. The majority of research, however, that has evaluated the merit of this unidimensionality concludes that the measure is not uni- but *multi*-dimensional.

Several two-factor solutions have been proposed for the GHQ-12 using either common factor or PCA approaches in conjunction with CFA. Most of these studies find some variant of "Depression/Anxiety" and "Social Dysfunction" constructs. The "Depression/Anxiety" construct tends to be given by items 2, 5, 6, 9, 10 and 11 (as listed in Table 2.1) and relates to the emotional side of psychological distress – with items concerning issues such as feeling unhappy, losing confidence, feeling under strain or worrying incessantly. The "Social Dysfunction" construct tends to relate to items 1, 3, 4, 6, 7, 8 and 12, that is more pragmatic notions such as the ability to carry out day to day functioning, issues with decision-making, lack of enjoyment of activities and struggling to face up to problems. This structure, although sometimes with different labelling , has at various points been identified as the best fit in data `from the UK (Smith et al. 2010), New Zealand (Kalliath, O'Driscoll and Brough, 2004), Brazil (Gouveia *et al.*, 2010), Japan (Doi and Minowa, 2003; Suzuki *et al.*, 2011), Germany (Schmitz, Kruse and Tress, 1999), Italy (Politi, Piccinelli and Wilkinson, 1994) and Turkey (Kiliç *et al.*, 1997). Meta-analyses consistently find these two factors are the most commonly identified both in two- and indeed in three-factor solutions (Werneke *et al.*, 2000; Picardi, Abeni and Pasquini, 2001).

The additional third construct most commonly refers to some form of "Loss of Confidence". This structure was initially found in the seminal paper by Worsley and Gribbin (1977) who carried out the first factor analytic study of the GHQ-12. They found the "Loss of Confidence" construct alongside "Social Performance" and "Anhedonia-Sleep Disturbance" which approximate the "Social Dysfunction" and "Depression/Anxiety" factors respectively. For Worsley and Gribbin this third posited dimension was constructed of four items (6, 9, 10 and

11) all pertaining to feeling worthless or depressed, or losing confidence, and has been supported by a number of subsequent works (Campbell, Walker and Farrell, 2003; Vanheule and Bogaerts, 2005; Penninkilampi-Kerola, Miettunen and Ebeling, 2006). Notably, Worsley and Gribbin did not subsequently refine to a simple structure solution, as there were no other structures to test against, so their proposed solution contains non-zero cross loadings. A very similar and arguably even more influential structure was published by Graetz in 1991. Graetz suggested that due to the GHQ-12 being a refinement of longer versions of the GHQ, which have consistently been shown to be multi-dimensional, it is likely that this shorter version is too (Martin, 1999). Graetz analysed a large scale (N= 8998) Australian sample, resulting in a three-factor structure comprising “Anxiety/Depression”, “Social Dysfunction” and “Loss of Confidence”. The Loss of Confidence factor proposed is very similar to that of Worsley and Gribbin – but is presented in the simple structure format of CFA, meaning that non-zero cross-loadings have been eliminated. Graetz’s Loss of Confidence factor only loads on items 10 and 11, which are the only two of Worsley and Gribbin’s factor that do not also load on another factor. This 3-factor structure has been the most widely accepted multi-dimensional structure since its introduction, supported by subsequent work both in the UK (e.g. Cheung, 2002; Martin & Newell, 2005; Shevlin & Adamson, 2005) as well as elsewhere (e.g. Gao et al., 2004; Padrón et al., 2012; Sánchez-López & Dresch, 2008).

It is clear from this review that there are a number of empirically and substantively different factor solutions proposed in the literature. Table 2.2 presents seven exemplar studies that characterise the range of structures that have been identified alongside the original unidimensional interpretation, detailing when they were published, and a brief outline of the factor structures they advanced. The table is organised by increasing complexity and highlights how since the original Goldberg formulation, proposed structures have tended to become more complex with time. All of these studies have been developed using traditional factor analysis techniques, comprising combinations of EFA (via common factors or principal components) and CFA in some form. In the next section ESEM is used to explore the structure in a less restrictive manner. This involves the relaxation of cross-loadings, and subsequently a four-factor solution is derived. This structure is then tested empirically against the seven predecessors to ascertain which fits the data better and ascertain which has the better fit using traditional factor analysis metrics. This chapter then goes further and evaluates these structures in both traditional and Bayesian estimation and present fit statistics for both. Ultimately this allows us to identify with a large degree of confidence the solution which most fully captures the variation in individual GHQ-12 likert scores for a large, representative sample from the UK.

Authors	*Date*	*Factor Structure Details*
Initial GHQ-12 Goldberg Formulation	1972	1 Factor – Baseline unidimensional model, all items specified to load on a single factor.
Hankins	2008	1 Factor – Unidimensional but with error covariance specified on negatively phrased items.
Andrich & Van Schoenbroeck	1989	2 Factor – Split into positive and negative items, each constituting a separate dimension.
Kilic et al.	1997	2 Factor – Anxiety/Depression, Social Dysfunction
Worsley and Gribbin	1977	3 Factor – Anhedonia-Sleep Disturbance, Social Performance, Loss of Confidence, specifies cross loadings on items 1 (F1, F2), 6 (F1, F3), 9 (F1, F3) and 12 (F1, F2)
Graetz	1991	3 Factor – Anxiety/Depression, Social Dysfunction, Loss of Confidence
Sanchez-Lopez & Dresch	2008	3 Factor – Successful Coping, Self-Esteem, Stress, Includes one non-zero cross-loading on Item 9 for F2 and F3.

Table 2.2: Seven exemplar studies of the range of factor analytical structures obtained from GHQ-12 data.

2.2 Methods:

2.2.1 Data

The data used here comes from a single source – the General Population Sample of the first Wave of the Understanding Society Dataset (US). Understanding Society (formerly the British Household Panel Survey - BHPS) is a large-scale, nationally representative annual panel survey of over 43,000 UK respondents from over 26,000 households (McFall, 2011). The study design incorporated and overhauled the pre-existing BHPS, extending its clustered, stratified postcode sampling to increase the panel size fivefold. Of these respondents 40,452 completed the GHQ-12, which was collected via face-to-face interviews. Likert responses were recorded for each of the 12 items; this allows ordinal modelling without the loss of information that would result from collapsing to binary outcomes.

2.2.2 Factor Analysis

Contemporary Factor Analysis techniques, partition variance into "shared" variance and "unique" variance. This can be done using so-called sufficient statistics, which work by reducing the full initial individual data to a correlation matrix and performing procedures on this, seeking to recreate this matrix using the smallest number of factors. Once an initial factor structure has been identified, the solution is rotated through multidimensional space to give the best diagnostic capacity to each of the underpinning structures. This involves changing the hypothetical end points of each of the latent axes. This aims to produce an interpretation where items load more strongly on fewer variables to give the researcher greater power in distinguishing the driving mechanisms underpinning them (Sass & Schmitt, 2010). The interpretation of the rotated solution is therefore often strongly influenced by which method is used to rotate the factor solution. This is another methodological choice in Factor Analysis that is often not given due consideration (Fabrigar et al. 1999).

2.2.3 Rotation Criteria

In order to evaluate the proposed solution, it is rotated through multidimensional space in an attempt to simplify its interpretation. That is, the specification of the solution has infinite possible orientations within the multidimensional space that will explain the data equally well. Factor rotation aims to aid interpretation of the solution by identifying structures most simply interpretable in terms of relating clearly to constituent item questions. This is operationalised by rotating the structure to try and produce larger loadings on fewer items for each constituent question (Thurstone, 1947, Fabrigar et al. 1999).

Factor Analysis has seen further improvement with the innovation of “oblique” rotation methods. This allows the latent constructs to be realistically modelled as correlated rather than specifying the explicit orthogonality necessary for earlier rotation methods. (Fabrigar et al. 1999). Geomin rotation is used here, a form of oblique rotation based on work by Yates (1987). It is the optimal method for obtaining clearly defined factor structures when little is known about the true underlying structure, and when variable complexity is suspected to be greater than one; that is to say, when variables load on more than one underpinning factor (Browne, 2001, McDonald, 2005, Asparouhov & Muthen, 2009). Whilst the developers of the rotation method caution its use when very complex structures are proposed, this is suggested to only affect cases where there are more than 4 factors, or variable complexity exceeds 3, which is not an issue for our proposed solution (Asparouhov & Muthen, 2009).

2.2.4 Factor Retention Criteria

In traditional factor analysis, when considering the number of factors to retain, model selection has been most optimally carried out by a process called Parallel Analysis in which a series of randomly generated "parallel" data sets mirroring the observed data set in sample size and number of explanatory variables (Hayton et al. 2004). This is used to determine that the factors observed and estimated are real, non-random, and generate larger eigenvalues than those from the randomly generated parallel estimation (Schmitt, 2011, Horn, 1965). More simple techniques (commonly set as defaults in statistical packages) include cut-off criteria for estimated factor eigenvalues, such as the Kaiser criterion denoting the retention of factors with eigenvalues greater than one (Kaiser, 1960, Hayton et al. 2004). However, this chapter deals with categorical data, for which parallel analysis has been shown to be problematic (Debelak & Tran, 2016, Muthen & Muthen, 2016, Weng & Cheng, 2005, Van der Eijk & Rose, 2015). Factor selection in the more advanced ESEM relies on comparing model fit statistics of different exploratory factor structures (Asparouhov & Muthen, 2009). Thus, fit statistics for initial structures with between 2 and 5 factors are presented here, illustrating which model best fits the data before going on to substantively interrogate the implications of this structure.

2.2.5 Exploratory Structural Equation Modelling (ESEM): Model-building Procedure

In 2009, Asparouhov and Muthen proposed Exploratory Structural Equation Modelling as a method by which to relax some of the more restrictive assumptions of CFA. ESEM has been suggested to represent a combination of the best elements of both the restrictive CFA; and structure-free Exploratory Factor Analysis (EFA) (Marsh *et al.*, 2009). Thus, given the subject matter, this new ESEM approach is adopted within the Mplus software environment (Muthén and Muthén, 1998-2014). The *key* contribution of ESEM methodology for this research is the specification of non-zero cross-loadings on constituent items (Asparouhov and Muthen, 2009). As outlined above, this is also crucial for realistic (not overinflated) estimation of factor correlations (Marsh *et al.*, 2010, 2014; Asparouhov, Muthén and Morin, 2015).

Given the strong arguments against oversimplification in psychological studies, and the empirical shortcomings of traditional Factor Analysis, an alternative step was taken after the initial EFA (using the common-factors approach and Geomin rotation), by not refining to simple structure. Indeed, instead of eliminating all but the largest loadings, *all* loadings that were greater than 0.15 were preserved in the EFA solution, and then the confirmatory element of the ESEM was rerun with these[7]. As the need to exclude loadings purely for methodological reasons has been relaxed, the value of 0.15 is a far more tolerant exclusion criterion than typically advocated in the literature (e.g. Ford et al. 1986, Huppert et al. 1989, Padron et al. 2012). This results in the proposed solution which shows that 8 of the 12 items load on more than one underpinning dimension. As seen below, this offers both empirically and substantively superior interpretation of the GHQ-12 structure.

2.2.6 Model Specification and Estimation

The GHQ-12 responses in the model are specified in the model as ordered categorical (ordinal) variables. For responses such as these the default estimator in Mplus is a robust weighted least squares estimator (Muthén & Muthén, 1998-2011). This estimation specifies a series of probit regression equations for each factor indicator (here questionnaire items) on the related factor (Muthén & Muthén, 1998-2011, pp. 62). These can be interpreted as likelihood changes in the

[7]. If any of the retained loadings fell below this threshold having rerun the model, they were eliminated, and the model rerun, however this only occurred for a single loading, and its removal did not affect model fit.

log-odds of changing response category on the response variable, illustrating the strength of the relation between the underpinning dimension and the probability of response change in the associated item (Chen, 2007).

All specified models are estimated using traditional Maximum Likelihood (ML) and the newer Bayesian estimation. Under the ML specification, variances of the factors are assumed to be normally distributed, with a variance of 1 and a mean of zero. The distribution of variance estimates is highly likely to be non-normal due to the constrained nature of variances and correlations, with variances having a lower limit of zero, and correlations being bounded by -1 and 1. Bayesian estimation does not require this assumption of normality of the parameter estimates. Under the Bayesian specification, prior variance-covariance estimates are drawn from an inverse Wishart distribution (Asparouhov and Muthén, 2010). The Wishart distribution is a multivariate extension of the Gamma distribution, allowing potentially skewed estimates for each dimension. Furthermore, Bayes estimation offers full uncertainty modelling, acknowledging explicitly the modelled nature of estimated parameters at each stage of estimation, allowing full error propagation at each iteration.

Results are presented as standardised loadings specified in the ESEM, which are interpreted as probit regression coefficients. This standardisation is crucial as due to the latent nature of the underpinning dimensions there is no inherent scale to the results. Indeed, without standardising the results by imposing at least one constraint there would be no way to identify the model units. The normal practice of this standardisation is carried out here by constraining factor variances to 1 (Muthén and Asparouhov, 2012). This means that item loadings (as probit regression coefficients) are interpreted on this scale and are thus directly comparable.

2.2.7 Evaluation of Fit

Having developed a new factor structure using ESEM, it needs to be compared with previous factor structures to see whether it fits the data better than the alternatives identified in Table 2.2. A a series of model fit statistics are estimated and presented. ESEM produces comparable fit statistics to traditional CFA measures which can be used to evaluate the structure (Marsh et al. 2014). The results for a selection of these fit indices are provided, each of which have associated strengths and weaknesses which are detailed below.

Root Mean Square Error of Approximation (RMSEA) is a maximum-likelihood based, absolute fit index, meaning that it does not compare with a reference model, simply comparing with a fully saturated model that perfectly reproduces the data. For the RMSEA lower values indicate better fit, with values below 0.05 tend to be considered indicative of “acceptable fit”. Conversely, the *Comparative Fit Index* (CFI) and *Tucker-Lewis Index* (TLI) are incremental fit indices, meaning that they measure the improvement in fit of the target model relative to a more restrictive, simpler model, usually one in which all observed variables are specified uncorrelated (Hu & Bentler, 1999). Of the two the TLI penalises more strongly for model complexity. The CFI is bounded by 0 and 1, with values closer to 1 indicating better fit. The TLI does not have this upper bound, but in practice rarely exceeds 1. For both CFI and TLI values over 0.95 are considered indicative of good fit. *Weighted Root Mean Square Residual* (WRMR) values are also presented. WRMR is calculated instead of the more traditional *Standardised Root Mean Square Residual* (SRMR) when response variables are categorical. Due to its relatively recent proposal (Muthen and Muthen, 2001-2014) there is considerably less literature regarding acceptable thresholds. WRMR itself is a residual-based measure, reflecting the weighted differences between the sample variance-covariance matrix, and the estimated population variance-covariance matrix. The cut-off bounds suggested (<1.00) for its use are associated with relatively small sample sizes (N=100, 250, 500) (Yu, 2002). As the calculation for categorical variables scales with sample size divided by sample statistics, it is likely to be upwards biased in cases with large sample sizes – here 400 times larger than Yu's validation study, so it is instead interpreted as a relative fit statistic compared with the simpler models. Values for CFI, TLI, RMSEA and WRMR are provided for all proposed structures as outlined in Table 2.2 in accordance with guidelines from Bentler (2007)[8]. It should be noted, however, that these guidelines were presented at a time where SEM was mainly undertaken in educational or psychological studies with small sample sizes. When dealing with a sample of the magnitude considered here, it is necessary to be more stringent with the selection criteria as all models can be expected to meet the sufficient fit statistic criteria (Marsh, Hau and Wen, 2004).

The Bayesian estimation allows a different yet complementary approach to evaluating the model. In this Bayesian framework, it is important and appropriate to make explicit a priori that every model is wrong, it is fundamentally a simplification of the raw data, and thus the model is unlikely to achieve a perfect predictive capacity. For the model runs using Bayesian estimation; each is specified for 100,000 iterations with uninformative priors, and posterior predictive checking (PPC) is carried out and presented via posterior predictive p-values. These are taken to be an indication of the model's capacity to reproduce the data and summarise the posterior distribution of the residuals (Asparouhov and Muthen, 2010). PPC is calculated as a likelihood ratio Chi-Squared statistic between the observed sample and a synthetic set of observations drawn from the posterior distribution. As such, it would be expected to see very low p-values, and associated positive confidence intervals which do not span zero. This is due

[8] For a more detailed discussion of specific guidelines and algebraic formula of each index see Hu & Bentler, (1999) and Schmitt (2011).

to PPC being based on chi-square and thus also experiences power increase with sample size (Muthen & Asparouhov, 2012). The larger the p-values, the less extreme the predicted values would be relative to the observed data and it is not realistic to assume the capacity to perfectly reproduce the mental health experience of some 40,000 individuals. However, whilst p-values are not expected to not differ significantly between models, the PPC also produces a 95% confidence interval for the discrepancy between the Chi-square model test statistic for the *observed* sample data and the *predicted* data (Muthen, 2010). This 95% posterior predictive confidence interval are more diagnostic of the capacity of the model to reproduce the data, where *absolute* replication of the observed data is unrealistic or unlikely, in our case due to a very large sample size of a heterogeneous UK-wide population (Marsh et al. 2004, Gelman, 2007). Thus it is assumed that reduction towards zero of the predictive confidence interval indicates an improvement in fit, and that the coverage range closest to zero will indicate the model with the best predictive capacity (Gelman *et al.*, 1996). The research simply aims to suggest a *better* model, or the best of the tested models. This chapter thus presents both the Posterior predictive p-value and the upper and lower bounds of the predictive 95% confidence intervals alongside the traditional, frequentist metrics. Additionally, when there is conflict between the results of the Maximum Likelihood estimation results and the Bayesian results, it will prioritise the Bayesian outcome as it derived from a less restrictive, and more realistically complex, estimation procedure.

2.3 Results

2.3.1 The Proposed Four-Factor ESEM Solution

Identifying the best factor structure is not simply the application of a statistical procedure, there must be theory underpinning the decision-making process. As such the proposed four-factor solution is first reported alongside the other 2, 3 and 5 factor-solutions given by the initial EFA. This section then provides details of the full four-factor structure, along with their modelled correlations to ascertain substantive and empirical dissimilarity of the resultant factors. Finally, it goes on to evaluate the proposed four factor structure empirically against the seven multi-dimensional factor structures tabulated earlier. Maximum-Likelihood and Bayesian fit statistics are presented for each structure and mean absolute factor correlations are also provided. These factor correlations are provided as mean absolute values as the magnitude, not the sign, of the correlation is of true interest, and it provides a convenient estimate of the substantive dissimilarity of the estimated factors of a model.

Table 2.3 shows the statistical improvement offered by a 4-factor model from the unidimensional baseline, giving Posterior Predictive P-values and predictive confidence intervals as model improvement between 2, 3, 4 and 5 factor solutions. In the traditional Maximum Likelihood Estimation, there is evidence for better model fit with added factors all the way up to a five-factor solution, with all four goodness-of-fit statistics improving with each added dimension. In the more empirically sound Bayesian estimation, it is clear that none of the specifications provide an adequate fit to accurately reproduce the 40,000 individuals' mental health responses, as evidenced by the p-values of 0.000. There is still evidence in the improvement of fit offered by adding factors in the predictive confidence interval. However, the improvement in fit associated with increased dimensionality evidenced in the Maximum likelihood estimation is here accompanied by a caveat. The improvement in predictive capacity is evidenced until the four-factor solution, however the five-factor solution is far less robustly estimated, evidenced by the very large confidence bound surrounding the Posterior Predictive Capacity. There is a small chance that the addition of the fifth factor *could* serve to improve predictive capacity as evidenced by the slight improvement in the lower bound of the confidence interval. However, there is a far greater chance that it would reduce the predictive capacity of the model evidenced by the very large upper bound, greater even than that of the three-factor solution, reflecting a great deal of uncertainty in the capacity of the model to replicate observed values. The four-factor model clearly offers the best predictive capacity, as evidenced by the lowest confidence interval. Therefore, it is the four-factor structure that is carried forward as it offers the most robust reproduction of observed values and, as we shall see, the most theoretically substantive factors.

EFA	*Maximum Likelihood*				*Bayesian*		
Number of Factors	CFI	TLI	RMSEA	SRMR	Posterior P-Value	2.5% CI	97.5% CI
2	0.975	0.962	0.086	0.035	0.000	2649.177	2987.687
3	0.993	0.986	0.052	0.017	0.000	459.431	624.921
4	0.997	0.992	0.040	0.011	0.000	195.417	334.959
5	0.999	0.995	0.031	0.007	0.000	59.243	919.419

Table 2.3: Bayesian and ML Fit statistics for EFA factor solutions for the GHQ-12 with 2, 3, 4 and 5 factors.

Having identified that the four-factor solution was the most appropriate from the estimation of the initial EFA whilst all cross-loadings are specified and estimated, the subsequent solution needed refinement to a testable solution. This was carried out by re-specifying the model having omitted all cross-loading values under 0.15. It is presented below in Table 2.4, with Factors labelled “Lowered Self Worth”, “Social Dysfunction”, “Stress” and “Emotional Coping”. These labels were chosen in accordance with existing names for previously identified factors from the literature. The loadings are standardised probit regression coefficients and are directly comparable both within and between factors. The values in parentheses are the associated standard errors for each factor loading. The blank spaces represent variables which were not specified to load on that factor as their loading in the initial EFA was low enough to warrant exclusion (<0.15).

GHQ-12 Items	F1 Lowered Self Worth	F2 Social Dysfunction	F3 Stress	F4 Emotional Coping
1. Able to Concentrate		0.627 (0.010)	0.229 (0.009)	
2. Loss of Sleep	0.548 (0.022)		0.464 (0.007)	
3. Playing a useful part		0.619 (0.005)		
4. Capable of decisions		0.856 (0.015)		-0.233 (0.017)
5. Constantly under strain	0.548 (0.028)		0.596 (0.008)	
6. Problem overcoming difficulties	0.656 (0.016)		0.339 (0.007)	
7. Enjoy day-to-day activities		0.554 (0.011)	0.198 (0.009)	0.176 (0.012)
8. Ability to face problems		0.732 (0.004)		
9. Feel unhappy/depressed	0.684 (0.014)		0.269 (0.007)	0.190 (0.010)
10. Losing confidence	0.924 (0.002)			
11. Think of self as worthless	0.854 (0.003)			
12. Feeling reasonably happy		0.514 (0.020)		0.469 (0.017)

Table 2.4: Standardised Factor Loadings for 4-Factor ESEM Model of the GHQ-12. Standard Errors in brackets. Calculations performed in Mplus using data available at https://www.understandingsociety.ac.uk/documentation/data-releases

Table 2.4 gives the loadings for the rotated solution as found via the ESEM approach along with the modelled correlations between the factors. The main difference from most of previous structures, is nonzero cross-loadings with items 1, 2, 4, 5, 6, 7, 9 and 12 all having associated cross-loadings, so that they are not simply the product of a single underlying dimension. The loadings of Factors 1 and 2 are consistent with previous literature, and thus have been given similar if slightly adapted names to reflect this. They differ empirically of course because they contain cross-loading items, but the broad differentiation between the positive and negatively worded items holds true. Higher factor scores for individuals on either of these items indicates higher levels of mental distress. As outlined earlier, associations between these items are not simply the result of wording effects, this oversimplifies the matter. Whilst they may be worded differently – there was an initial thought process which has grouped positive and negative aspects of mental health, and the GHQ-12 tried to capture both. Trying to homogenise the meaning of these dimensions by presenting them solely as an artefact of the wording seems reductive. In addition, the assertions of Marsh et al. (2014) of the artificial inflation of correlations in models imposing simple structure are borne out clearly in the correlations between these modelled factors in Table 2.5. Despite them being composed very similarly to those in the previous literature, the correlations found here are far lower due than commonly reported.

The third factor, here termed “Stress” is seen to load most strongly on items to do with feeling under strain and loss of sleep (Items 5 & 2 respectively). It is in fact very similar to a factor found by one of the earliest factor analyses of the GHQ-12 by Worsley and Gribbin (1977), differing only in that they found it also loaded on item 12. They termed this “Anhedonia – Sleep Disturbance”. This factor is not commonly seen outside of this early paper, which is very likely due to it requiring cross-loadings to be specified, as they did not refine to simple structure to test against other solutions. Beyond this initial suggestion of the factor, there is little mention of a structure analogous to this, other than in a “Thematic Analysis” of the GHQ-12 by Martin in 1999. Coming from an educational psychology background, he broke from tradition, choosing to look at the items based on content, rather than exploratory statistics, to suggest a factor structure, which he then tested empirically and found fitted the GHQ-12 better than both the Graetz (1991) and Worsley and Gribbin (1977) solutions. He identified both “Stress” and “Successful Coping” factors. The suggested “Stress” factor also primarily loaded on Items 2 and 5, as with the Stress factor here. These results are entirely in line with the four-factor solution here, given that the elimination of the smaller cross-loadings would have led us to the same result were it not possible to relax the constraints associated with CFA.

The emergence of the fourth factor, termed "Emotional Coping" is a distinctive finding and it is interesting and important for several reasons. It is *negatively* associated with Item 4 – "feeling capable of making decisions", but *positively* associated with feeling unhappy or depressed, not enjoying day-to-day activities and not feeling happy (Items 7, 9 & 12 respectively). It is worth reiterating that all raw items scores are coded at the outset such that higher scores always indicate greater mental distress[9]. Emotional Coping is structurally similar to the "Loss of Confidence" construct found in work by Graetz (1991) with one key difference. Emotional Coping is notable for having negative coefficient for a cross-loading, where the negative loading is on Item 4, an item clearly associated with positive coping. The factor is most strongly given by negative emotion, but clearly has a negative association with decision making. This seems to suggest that those individuals evidencing highest distress as captured by Emotional Coping are those who also consider themselves most capable of decision making, they persevere and continue to evaluate themselves as having uncompromised decision-making functionality in the face of this distress. It is interesting that the items capturing this negativity are phrased both positively and negatively suggesting that phrasing does not matter for this construct. For this reason, it is named "Emotional Coping". The Emotional Coping factor has the unique ability to mask the expression of the other factors in a composite score due to its negative loading. Substantively this seems plausible, it is not hard to envision an example of a

[9] It is also important to remember that as a latent variable, an equally valid interpretation is the exact inverse, with negative loadings on Items 7, 9 and 12 and a positive loading on Item 4, which would in turn invert the poles of the underlying coping dimension, with higher scores indicating vulnerability rather than tolerance.

high scoring individual on this factor; someone traditionally very stoic, their evaluation of their decision-making capacity unaffected by their very negative emotional experience. They could be experiencing very low self-worth, and very high levels of depressive symptoms, but this be masked to a degree in their aggregate GHQ-12 score because they are still very good at "coping". They may, for example, still be able to carry out normal social functioning, borne out in Factor 2, and yet be experiencing deep internal distress. Conversely, a low scoring individual would be far more cautious in their evaluation of their decision-making capability, regardless of their current state of negative emotion. The complexity of this factor presents another strong argument for a more nuanced understanding of mental health as captured by the GHQ-12 as this intricacy would be missed by simple summation which presumes that each and every item captures mental health in the same 'direction' and on the same dimension. Whilst the Emotional Coping factor suggested here is unusual in GHQ-12 literature, it is not unique. The existence of a "Successful Coping" factor underlying the GHQ-12 was also suggested by Martin in 1999. Moreover, this is not the first notion of coping as a mechanism influencing the expression of psychological symptoms, as this existed in the psychological literature for decades (e.g. Collins et al., 1983; Folkman et al., 1986; Petrosky & Birkimer, 1991). In recent years, several other GHQ-12 factor analysis papers have proposed factors with labels indicating some element of "coping", such as the Sanchez-Lopez & Dresch (2008) structure. Crucially, it is a factor which would not have been possible to identify using traditional Factor Analysis techniques, as it is constructed entirely of items with cross-loadings.

Having detailed the theoretical implications of the proposed factor structure it is necessary to understand the empirical and substantive implications of the structure. This is done by evaluating the substantive dissimilarity between factor constructs as measured by the correlations between those constructs. These are presented in Table 2.5.

FACTOR CORRELATIONS	F1 Lowered Self-Worth	F2 Social Dysfunction	F3 Stress	F4 Emotional Coping
F1. Lowered Self Worth	1.000			
F2. Social Dysfunction	0.680	1.000		
F3. Stress	0.178	0.111	1.000	
F4. Emotional Coping	0.411	0.487	0.271	1.000

Table 2.5: Modelled Factor Correlations from the Four Factor ESEM Solution for the GHQ-12

Modelled factor correlations are an important indicator of the capacity for discrimination between the proposed factors. If the factors are very similar with a high correlation, they may not offer much substantive inference over a unidimensional structure and pursuit of parsimony would, and has in the past, lead the researcher to treat the constructs as unidimensional. However, such high correlations may be an artefact of refining to simple structure. Here this is borne out very clearly in the far lower factor correlations than are typically reported in the literature. This suggests that the factors are more substantively dissimilar than those traditionally suggested in the literature. There are clearly different structures underpinning the GHQ-12, echoing our earlier assertion that there is far more going on than could be captured by traditional interpretation with fewer dimensions[10].

The greatest similarity is between the Lowered Self-Worth and Social Dysfunction factors, although this correlation is notably low relative to that found in the existing literature. This seems to sit well with the literature being contested here which has found only these two factors

[10] It is worth highlighting that factor correlations overestimate the predictive capacity of each variable on the other. The *absolute* proportion of variation in one variable that could be predicted solely from knowing the other is given by the squared value of the correlation coefficients (Kish, 1954). For example, knowing the modelled Lowered-Self-Worth scores for all individuals would only allow the prediction of 46.24% (0.4624) of the variation in Social Dysfunction scores, despite these factors having the highest modelled correlation of 0.68. This is even more stark for the correlation between Stress and Social Dysfunction, with a true predictive capacity of 1.2%, leaving 98.8% of variation in one unexplained by the other.

and often concluded due to their similarity that they are essentially capturing the same thing. Stress is the most statistically dissimilar and therefore most substantively distinct factor. Those experiencing high levels of stress do not necessarily report lower self-worth or lowered social functioning, nor are they necessarily more stoic. Furthermore, despite the negative loading, the "Emotional Coping" factor is positively correlated with the other 3 factors. This suggests that individuals who are experiencing more distress as captured by any of the metrics are more likely to be stoic as captured by this Emotional Coping metric. The most notable characteristic of the solution is the low correlations between the four constructs relative to existing findings, suggesting that there are clearly distinct processes going on within what is being captured by the GHQ-12.

The four-factor solution holds up to substantive scrutiny. It sits well within the literature, and the factors are more substantively and empirically distinct than found in other solutions. All that remains is to test against other structures on the same dataset to verify its improvement. against alternatives.

2.3.2 Confirmatory ESEM – Evaluating the Four Factor Structure against Alternatives

Table 2.6 gives a detailed comparison identifying the factor loadings estimated under each specification. The comparison models are among the most heavily referenced in the literature, alongside the often-used unidimensional control – as well as the proposed ESEM solution in the final column. The table demonstrates the differences in structure and complexity between proposed GHQ-12 solutions. It also clearly shows the greater complexity of the four-factor solution.

GHQ Item	Uni	Hankins (2008)	Kilic et al. (1997)		Andrich & Van Schoubroeck (1989)		Sanchez-Lopez & Dresch (2008)			Graetz (1991)			Worsley and Gribbin (1977)			ESEM (2017)			
Factor #	1	1	1	2	1	2	1	2	3	1	2	3	1	2	3	1	2	3	4
1	x	x		x	x		x			x			x	x		x		x	
2	x	x!	x			x			x		x		x				x	x	
3	x	x		x	x		x			x				x		x			
4	x	x	x		x		x			x				x		x			x
5	x	x!	x			x			x		x		x				x	x	
6	x	x!	x			x		x			x		x		x		x	x	
7	x	x		x	x		x			x			x	x		x		x	x
8	x	x		x	x		x			x				x		x			
9	x	x!	x			x		x	x		x		x		x		x	x	x
10	x	x!	x			x		x				x			x		x		
11	x	x!	x			x		x				x			x		x		
12	x	x		x	x		x			x			x	x		x			x

Table 2.6: Factor Structure of all eight tested GHQ-12 solutions including new ESEM solution. "x" indicates a specified loading, "!" indicates error covariance. Positively worded Items emboldened, Negatively worded items italicised.

Table 2.7 presents fit statistics for all seven alternatives alongside the proposed ESEM solution. The mean absolute factor correlation is also presented for each specification to provide an estimate of the dissimilarity of the factors estimated in the model structure. The four-factor solution provides the best fit by a considerable margin across each and every measure of fit. Moreover, this is the case under both traditional and Bayesian estimation. There is notably little improvement between the unidimensional specification and the initial Kilic et al (1997) specification, the secondary factor does not seem to make the model fit the data any better. Also notable is that the model proposed by Hankins (2008) with covariance specified on negatively worded items performs very well compared with other 1- and 2- factor solutions.

The best performing structure outside of the ESEM solution is the Worsley and Gribbin (1977) structure with its non-zero cross-loadings. Factor structures which address the inter-dependencies of the items via error covariance or non-zero cross-loadings perform very well across all the fit statistics, which again is an argument for the adoption of a more realistically complex specification of mental health. Hankins (2008) suggested that all multidimensional factor structures were a product of over-interpretation of spurious variance in negatively worded items. This seems a reasonable assertion, but again only if one assumes that the questions were grouped into positive and negative items at random. As suggested earlier, this is not likely to have been the case in the design of the questionnaire, where questions are likely to have been grouped thematically. This would lead to similarities between responses over and above that which is solely a product of wording. This may present a problem in traditional factor analysis, however here these dependencies can be incorporated into the traditional factor analysis via the methodological advances offered by ESEM.

Factor Structure	No. of Factors	WRMR	CFI	TLI	RMSEA	Posterior predictive P	Lower Bound	Upper Bound	Mean Absolute Factor Correlation
Unidim-ensional	1	14.776	0.914	0.895	0.146	0.000	7123.436	7634.537	-
Hankins	1	6.808	0.981	0.967	0.082	0.000	1505.348	1764.743	-
Andrich & van Schoubroeck	2	9.018	0.967	0.959	0.091	0.000	3521.095	3892.429	0.777
Kilic et al.	2	12.826	0.934	0.918	0.129	0.000	5268.267	5709.397	0.822
Worsley & Gribbin	3	5.505	0.986	0.981	0.063	0.000	1199.430	1430.393	0.665
Sanchez-Lopez & Dresch	3	8.246	0.972	0.963	0.087	0.000	2824.202	3152.636	0.772
Graetz	3	7.917	0.974	0.966	0.083	0.000	2459.754	2769.927	0.783
ESEM 4Fac	4	3.247	0.995	0.992	0.039	0.000	376.571	526.539	0.356

Table 2.7: Bayesian and Maximum Likelihood Fit Statistics for the 8 evaluated factor structures defined in Table 2.6.

Note: Posterior Predictive P, Lower Bound and Upper Bound and Factor Correlations are all reported for Bayesian model specifications, whereas WRMR, CFI, TLI and RMSEA are all taken from the Maximum Likelihood Specification.

In the Bayesian results, as expected due to the sensitivity of PPC to sample size, *all* Posterior P Values are reported as zero to three decimals places. Despite this, differences in fit are clearly evidenced by the 95% PPC Confidence Interval and this supports the traditional fit statistics under the ML estimation. Although none of the coverage bounds span zero there is a marked difference between the model structures, with the four-factor solution clearly showing closest proximity to zero. This should be interpreted as clear evidence that the four-factor solution offers *better* but not *perfect* fit to the data even under the less restrictive and methodologically superior Bayesian specification, which lends more strength to the argument that it outperforms the other solutions.

The evidence of empirical improvement is further supported by the substantive improvement evidenced in the absolute mean correlations between the modelled factors under each of the seven alternative model specifications. The four-factor model has by far the lowest mean absolute factor correlation, suggesting the most dissimilar and empirically distinct factors. There is also clear supporting evidence across all fit statistics, but most notably in the modelled factor correlations, for the cautioning against over constrained simple structures by Asparouhov and Muthen (2009) and Marsh et al. (2014). The next best solutions are those which contain cross-loadings. The Sanchez-Lopez specification has only incrementally lower factor correlations on average than that in the Andrich and Schoubroeck specification. High correlation between factors is often cited as evidence for unidimensionality. This has, however, now been clearly demonstrated to be an artefact of the modelling process, and that constraining all cross-loadings to zero when items are truly correlated artificially inflates correlations.

2.4 Discussion

This chapter presents clear evidence for at least the re-evaluation of assertions of unidimensionality in previous literature with less constrained models to see if the conclusions are the same under more realistic assumptions. Furthermore, it is important to acknowledge that this is only possible because of the methodological innovation allowing this model specification, and that we are not attacking the work of the authors based on merit, rather on limited methodology at the time.

2.4.1 Modelling Mental Health: Realistic Complexity:

Despite the seemingly context-specific interpretation of the constituent items, it has been noted that traditional factor analyses across languages and settings generally find structures consistent with Graetz (Werneke et al. 2000). We find two typical structures, broadly categorised by positive and negatively phrased items, in line with much previous multidimensional literature. However, the additional factors we find are more in line with earlier work by Worsley and Gribbin (1977), who find similar constructs to the "Lowered Self-Worth" and "Stress" factors that we identify here, although we suggest this is likely a product of their being one of very few studies which reports non-zero cross-loadings on constituent items. Moreover, Martin (1999) and Sanchez-Lopez & Dresch (2008) both find evidence of a "Coping" dimension, and suggest that this fits the data better than that of Graetz. We find evidence of both in this work, with analogous factors to those proposed by Graetz being evidenced alongside an "Emotional Coping" factor. Whilst these factors are more complex than those proposed in the initial works, due to the specification of cross-loadings, both these structures are supported by this work. The resulting structure incorporates them both into a more realistic, encompassing four-factor solution, relaxing the constraints of typical Factor Analysis techniques from which these originally emerged.

The newly proposed structure offers empirical improvement over the previous seven structures across every single metric, as evidenced by fit statistics in Table 2.7. It offers more than just empirical improvement, however, as the factors themselves are more distinctly defined, as evidenced by the lower correlations in Table 2.5. This leads us to conclude that the factors identified here do not only explain the data better, but they are also more substantively distinct than those in previous work, which we suggest is strong evidence of several separate (although somewhat correlated) processes underpinning the GHQ-12 measure. We suggest that this is due to the greater methodological flexibility offered by the innovation of ESEM and the relaxation of the simple structure required in traditional CFA (Asparouhov and Muthen, 2009, Marsh et al. 2014).

2.5 Conclusions:

2.5.1 Moving Forward with Mental Health

The large-scale adoption of the GHQ-12 as a population screening tool has led to much research into its underpinning meanings. It is often suggested to be multidimensional, measuring more than the single 0-36 or 0-12 metric that it is so commonly used. as. Here we investigate ways in which to reinterpret GHQ-12 worldwide, in light of newer methodological advancement and understandings in mental health. We strongly advocate modelling procedures which explicitly acknowledge the nuanced and complex nature of a topic such as mental health, and caution against the traditional, simplistic and often reductive capacity in which GHQ-12 responses have been used

By utilising novel ESEM techniques, we conducted an exploratory investigation into the structure of the GHQ-12, finding 4 distinct underpinning dimensions. These dimensions are analogous to suggested features of the measure proposed in previous literature. Interestingly one of these structures, "Emotional Coping" has a negative factor loading, meaning that it could serve to mask the presence of psychological distress. We then tested this structure against the seven most highly cited interpretations from the GHQ-12 literature, estimating the model using both traditional frequentist and Bayesian estimation. We find that the solution is not only empirically superior, as evidenced by it performing better across every fit statistic, but that is also offers more substantive insight into the underpinning factors, as evidenced by the low absolute correlations. To this end, we suggest caution when refining to simpler structures based on factor correlations.

Substantively one of the most interesting things to come out of this paper is the emotional coping construct. There is a disconnect between the modelled at-risk demographics in self-rated mental health literature and those who are identified as more at risk with clinical outcomes. For example, males disproportionately die by suicide (Samaritans, 2016), whilst self-rated mental health studies largely conclude that women are experiencing more mental distress (e.g. Propper et al. 2005, Weich et al. 2003). It is speculative to guess at the reasons for this disconnect, however it does not seem unlikely that this is due to a difference in the self-evaluation of the individual to see themselves as at-risk. Whilst not modelling this directly, this quantifiable construct of Emotional Coping seems to represent the underpinning tendency of individuals to not see their negative emotional experience as affecting their capacity to perform functioning via decision making. More interestingly, the correlations associated with this factor seem to suggest that individuals are more stoic the more distress they experience, which has important implications for self-rated mental health interpretation more broadly. Whilst this is tentative, it signals a move towards a more realistic interpretation of complex and compound mental health measures, which signals *both* the individuals at risk *as well as* those who may not be detectable by the study even if they were.

Common across the results for the evaluated structures is a clear pattern favouring nuance in interpretation. Whilst it has been suggested that multidimensionality may a be simple product of mis-specified error covariance amongst negative items, we find that there is considerable variation *both* within and between the positive and negatively worded items. All models which allow a degree of inter-relatedness via error-covariances or cross-loadings outperform the simpler alternatives. This is evidenced not just in raw fit statistics, but also in the lower factor correlations suggesting distinct constructs removing the artificial inflation imposed by simple structure. The chapter demonstrates the need complexity in modelling mental health, which cannot simply be understood in aggregate simple summed scores

2.5.2 Recommendations for Further Study

The structure and associated underpinning constructs proposed here should clearly be considered when moving forward with GHQ-12 research as they offer substantive and empirical benefit over and above structures identified with more dated methods. Furthermore, the key differences between this newly proposed structure and previous structures are likely to be largely a product of the tendency of previous methodologies to posit too many restrictions on the interpretation of an inherently complex subject. To that end, the structure proposed here does not invalidate previous research, rather it validates and incorporates it into a more holistic understanding via the specification of newer and more realistic modelling approaches. It also offers information to researchers from both qualitative and quantitative backgrounds, in offering a new series of processes which are evidenced to be underpinning the GHQ-12.

To this end, standardised factor loadings are reported to allow composite score creation. The aim here is to allow future researchers to have a method by which they can combine GHQ-12 items into meaningful sub-scales based on the dimensions established by the ESEM analysis. This involves the treating of the standardised loadings as coefficients for the summing of responses in each factor across a population. These loadings could then be used to derive composite GHQ measures for use in longitudinal studies. Variance estimates based on these new terms should be interpreted with caution, however, as variance is likely to be inflated on negatively worded items as indicated in Figure 2.1 due to the lack of an obvious baseline category.

3 WHAT IS BEING TRULY COMPARED? RELATIONSHIPS BETWEEN MENTAL HEALTH, MENTAL ILLNESS AND WELLBEING

3.1 Background

3.1.1 Why care about positive mental health measurement?

A convincing case has been made for both the health and economic benefits associated with increased consideration and promotion of positive mental health worldwide (Howell, Kern and Lyubomirsky, 2007; Stiglitz, Sen and Fitoussi, 2010; Knapp, McDaid and Parsonage, 2011). High levels of positive mental health have been shown to be associated with better recovery from illness and reduced cardiovascular disease risk (Boehm *et al.*, 2012; Keyes and Simoes, 2012; Trompetter *et al.*, 2017), as well as being robustly associated with increased longevity, workplace and classroom productivity, and resistance to psychological decline with age at the population level (Beddington *et al.*, 2008; Huppert, 2009; Slade, 2010; Diener and Chan, 2011). As a result, positive mental health promotion is being explicitly incorporated into public policy and even growth forecasting in the UK alongside traditional mental health considerations (Department of Health, 2009, 2013; The Scottish Government, 2009; HM Government, 2012). Despite these acknowledged benefits, definitions of what is meant by "positive mental health" are often ambiguous, with "happiness" often simply evaluated by simplistic or single item responses (Weich *et al.*, 2011; Mukuria *et al.*, 2014). This has resulted in quantitative literature rarely considering mental health as more than "absence of diagnosable mental illness".

This increased acknowledgement of "positive mental health", often used interchangeably with "mental wellbeing", has been evidenced also in quantitative social science literature on the determinants of mental health (Oguz, Merad and Snape, 2013; Schrank *et al.*, 2013). However, despite "mental wellbeing" having ideological and conceptual roots beyond the simple categorisation of mental health into "ill" and "not", it is commonly operationalised in quantitative research as simply the absence of "mental illness" informed by the thematic design of early mental health monitoring questionnaires (Headey, Kelley and Wearing, 1993; Westerhof and Keyes, 2010). Whilst this sounds intuitive, a growing and robust body of evidence suggests that the processes governing positive and negative mental health differ, leading to calls for separate investigation of determinants and resultant health impact of each facet (Beddington *et al.*, 2008; Lamers *et al.*, 2015; Trompetter *et al.*, 2017). If indeed this duality holds true, it raises doubt over whether the application of lessons learned in traditional "mental illness" literature to the promotion of "mental wellbeing" is valid, or whether the determinants of wellbeing should be investigated as an entirely separate route of enquiry.

The evidence of the conflating of these terms in quantitative social science is commonly seen in the ascribing of "mental wellbeing" monitoring function, to questionnaires such as the GHQ-12 which were developed prior to the large-scale adoption of positive mental health (e.g. Tait, French and Hulse, 2003; Propper *et al.*, 2005). Several studies have questioned the appropriateness of inferring positive mental health outcomes from the GHQ-12, but as there is little consensus on what is meant by wellbeing in these studies, there is little agreement on whether this is valid practice (e.g. Huppert and Whittington, 2003; Kalliath, O'Driscoll and Brough, 2004; Hu *et al.*, 2007). This is especially pertinent as early mental health investigations largely tested for psychological distress and inferred mental health from its absence (Hollingshead and Redlich, 1958; Henderson, 1981), rather than explicitly addressing determinants of positive mental health. Given that policy recommendations and conclusions are still cited from this body of literature, it is important to be sure of what determinants and processes are being described or can be generalised to. More recently, mental health measures such as the Short Warwick Edinburgh Mental Wellbeing Scale (SWE) have been developed to explicitly capture positive mental health, allowing us to categorise more clearly any differences between positive and negative mental health determinants (Tennant *et al.*, 2007; Stewart-Brown *et al.*, 2009; Stewart-Brown, Samaraweera, Taggart, N. B. Kandala, *et al.*, 2015).

3.1.2 Moving Beyond Absence of Illness

There is a great wealth of literature in quantitative social science on the determinants of mental health however the approach to characterising “mental health” differs greatly between disciplines and authors. Typically, this does not move beyond the traditional “unidimensional” view of mental health outlined in Chapter 2, with the assumption that at the opposite end of the construct underpinning “mental illness”, lies an inverse but equivalent outcome of “wellbeing” (Headey, Kelley and Wearing, 1993). However, whilst conceptual literature has long acknowledged the possibility that positive and negative mental health outcomes may be driven by entirely different processes, the adoption of this perspective in quantitative social science has, with some notable exceptions (e.g. Weich *et al.*, 2011), somewhat lagged behind (Stewart-Brown, Samaraweera, Taggart, N. B. Kandala, *et al.*, 2015).

The separation of processes of positive and negative mental health has been proposed and evaluated at length in wellbeing literature (e.g. Massé *et al.*, 1998; Huppert and Whittington, 2003; Trompetter *et al.*, 2017). Just as causes of positive affect cannot be inferred from causes of negative affect, similarly mental illness cannot be thought of as simply the inverse of wellbeing (Cacioppo and Berntson, 1999; Ryan and Deci, 2001; Keyes, Dhingra and Simoes, 2010). Recent evidence has been accumulating in support of this “Two Continua” model of mental health (Greenspoon and Saklofske, 2001; Westerhof and Keyes, 2010; Lamers *et al.*, 2015). This approach does not suggest that positive mental health is unrelated to mental illness, however it does explicitly suggest that positive and negative mental health should be viewed as distinct and only moderately related (Westerhof and Keyes, 2010; Lamers *et al.*, 2015; Trompetter, de Kleine and Bohlmeijer, 2017). This perception of wellbeing as more than the absence of illness raises important questions about the processes and variables underpinning what is commonly understood by mental health.

The main issue that arises logically from this distinction is that treating mental illness and wellbeing as unidimensional and simply the inverse of each other is likely to be reductive. That is not to imply that they are unrelated; to the contrary, mental illness has been shown to longitudinally affect mental health (Zatzick *et al.*, 1997; Eack and Newhill, 2007) and positive mental health predicts reduction in psychopathological symptoms such as depression (Keyes, Dhingra and Simoes, 2010; Grant, Guille and Sen, 2013). The approach simply asserts that positive and negative mental health are related but, crucially, distinct. This is already accepted in psychological literature; for example, studies show childhood personality characteristics such as neuroticism are strongly and robustly predictive of later-life levels of mental distress, but absence of this is not predictive of wellbeing (Van Os, Park and Jones, 2001; Kendler *et al.*, 2005). Similarly, extraversion has been found to robustly predict positive emotional characteristics although again negative emotional experience is not necessarily predicted by the absence of extraversion (Clark, Watson and Mineka, 1994; Neeleman, Ormel and Bijl, 2001). Separating out the processes underpinning wellness and illness allows better investigation into the determinants and risk factors for each. Treating the predictors of illness as separate to those of wellness also potentially allows for greater interpretive complexity. This allows the more complex characterisation of certain at-risk groups, for example Westerhof and Keyes find that older adults displayed fewer symptoms of mental illness but no discernible difference in their experience of positive mental health than younger adults (2010). This echoes the call for increased nuance in mental health inference outlined in Chapter 2. Having established that the processes driving positive and negative aspects of mental health may be separate, a robust methodology for identifying these characteristics must be developed.

3.1.3 Capturing Positive Mental Health

Establishing that positive mental health constitutes more than simply the absence of illness proves far simpler than defining what positive mental health truly is. Mental wellbeing, like mental health more broadly, is a concept that is hard to define objectively. It relies on subjective self-scrutiny of a subject, on the basis that only the experiencing self is capable of evaluating their own happiness (Ryan and Deci, 2001). This is further complicated by the effect of social comparison, where the individual's *perception* of their context relative to a social hierarchy is often more important than their true position (Layard, 2006). Despite the obvious difficulties in identifying common themes, efforts have been made to capture what is meant by wellbeing. The WHO defines mental health as "a state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community", (World Health Organization, 2001, pp. 1). Psychological definitions tend to take on a more nuanced approach, recognising two distinct perspectives named "hedonic" and "eudaimonic" wellbeing. Hedonic wellbeing traditionally focuses on happiness and defines wellbeing in terms of pleasure-seeking and pain avoidance (Kahneman, 1999). Eudaimonic wellbeing focuses on individual self-realisation, and the perception of meaning and worthiness in the pursuits of the individual, viewing wellbeing in terms of realised potential (Waterman, 1993; Ryan and Deci, 2001). Although these elements are related, they cannot be treated as identical, with evidence suggesting that, for instance, whilst eudaimonia seems to somewhat predict hedonic wellbeing, it is not the sole predictor, nor does it tell the full story (Waterman, Schwartz and Conti, 2008).

Whilst in psychological literature this has been acknowledged for decades, this has not been translated into large scale survey questionnaires in the UK until recently. The Short Warwick Edinburgh Mental Wellbeing Scale (SWE) is a shortened version of a wellbeing questionnaire developed originally by Tennant et al. in 2007. It explicitly addresses both hedonic and eudaimonic elements of wellbeing by design, and is exclusively focused on the capturing and representation of positive mental health, considering both emotional experience and psychological functioning (Tennant *et al.*, 2007). Both the original 14-item questionnaire and the shortened 7-item version have been extensively validated using simple Rasch or confirmatory factor analytic methods in the UK (Stewart-Brown *et al.*, 2009; Bartram, Sinclair and Baldwin, 2013; Mukuria *et al.*, 2014) and overseas (Maheswaran *et al.*, 2012; Mater *et al.*, 2013; Castellví *et al.*, 2014). The scale has seen considerable uptake in the UK and has been included in several large scale surveys, including the Health Survey for England (HSE) and Understanding Society (US).

This chapter will evaluate the relationship between responses to the SWE questionnaire and the GHQ-12 responses that were seen in Chapter 2. The GHQ-12 is taken to represent "mental illness", given its design and development, and the SWE is taken to represent "mental wellness". Whilst some research has suggested the GHQ-12 can adequately capture positive mental health, the majority of evidence seems to suggest the opposite (Hu *et al.*, 2007; Davies, 2014). It is of key importance in understanding mental health to evaluate the contested capacity of negative mental health measures to capture positive mental health. Methodologically, the advocacy for increased nuance from Chapter 2 is echoed here, with the use of Exploratory Structural Equation Modelling (ESEM) to first unpack the SWE and then to investigate the degree to which it predicts or is predicted by the GHQ-12. This is carried out using the largest simultaneous distribution of these measures in the UK to date – comprising nearly 40,000 individuals taken from the first wave of Understanding Society, each responding contemporaneously to both survey instruments.

3.2 Data

The data used here is, again, taken from the General Population Sample (GPS) of the 2009 wave of the Understanding Society (US) Dataset. The 2009 wave was chosen as it was the first large-scale, UK-based survey to incorporate both the GHQ-12 and the SWE, obtaining responses to *both* surveys from 37,836 individuals, and a further 2616 who answered just one. These data were collected via face-to-face interviews. This represented the largest simultaneous distribution of these questionnaires in the UK, and indeed still does, with the exception of the 2011 wave.[11] The originally collected Likert responses are modelled here allowing us to model ordinality without collapsing to more traditional binary interpretation and thus losing information. This method is also consistent with the methodology outlined in Chapter 2 allowing for comparison of similarly constructed dimensions.

The SWE is comprised of 7 Items, each responded to on a Likert scale from 1-5, in a similar fashion to the GHQ-12. Responses are then summed to give a total score between 7 and 35, with higher scores indicating greater wellbeing. The Items are given below in Table 3.1:

[11] 2009 was selected over 2011 in order to try and minimise test-retest effects which are known to compromise external validity in large-scale psychological screening (Anastasi, 1976; Boyd, Le and Somberg, 2005).

Statements: *"I've been…"*	None of the Time	Rarely	Some of the Time	Often	All of the Time
1. Feeling Optimistic about the Future	1	2	3	4	5
2. Feeling Useful	1	2	3	4	5
3. Feeling Relaxed	1	2	3	4	5
4. Dealing with Problems well	1	2	3	4	5
5. Thinking Clearly	1	2	3	4	5
6. Feeling close to other people	1	2	3	4	5
7. Able to make up my own mind about things	1	2	3	4	5

Table 3.1: Items and Responses for the Short Warwick Edinburgh Mental Well-Being Scale

Figure 3.1 presents the raw responses to each item stratified by category. Responses across these categories are less sharply modal than the responses given to the GHQ-12 in Chapter 2. All items follow a similar response pattern, with the most notable deviation from this being Item 7 – "I've been able to make up my mind about things", where individuals far more commonly responded with "All the Time". Similarly, Items 1 and 3, pertaining to optimism and relaxation, have a greater number of individuals responding negatively or with "Some of the Time".

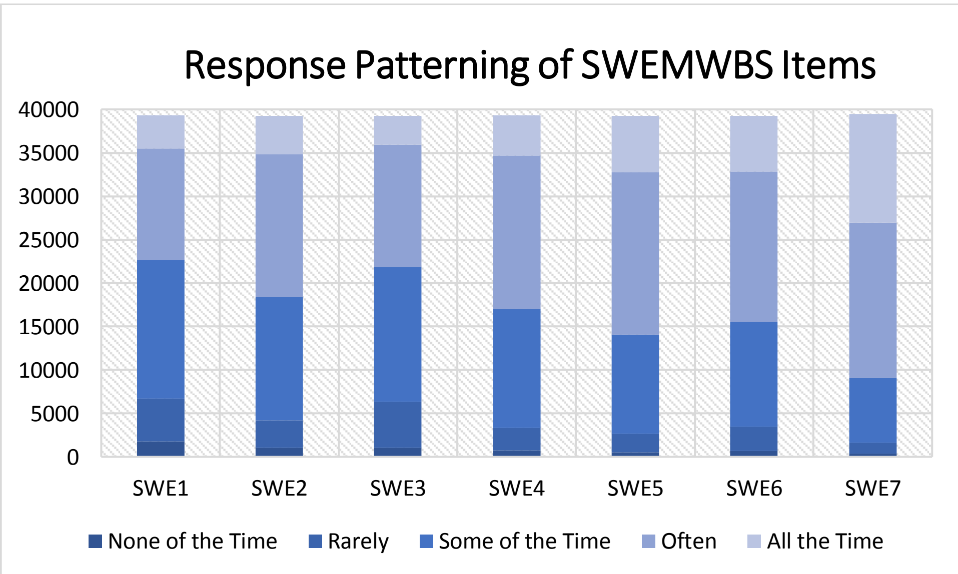


Figure 3.1: Observed Response Patterning for the SWE in 2009 wave of Understanding Society

3.2.1 Comparing Latent and Manifest Mental Health

Having characterised the SWE, an approach needs to be specified that can handle the methodological complexity of comparing the GHQ and the SWE. Furthermore, the modelling procedure needs to be extendable to the thematic and conceptual constructs underpinning the relationship between the observed responses.

There are two distinct epistemological viewpoints underpinning traditional quantitative analysis of questionnaire data. This ultimately refers to whether the outcome of interest is in the manifest observations, or in the underpinning latent constructs assumed to be dictating those manifest observations. When considering the relationship between the two scales the relationship between the observed or manifest questionnaire responses can be investigated, or the relationship between the latent constructs underpinning those questionnaires can be investigated.

Both perspectives are underpinned by a philosophical standpoint on measurement and objectivity. If the observed perspective is taken, as the overwhelming majority of previous research has done (Weich and Lewis, 1998; Propper *et al.*, 2005; Blanchflower and Oswald, 2008; Clarke *et al.*, 2011; Maheswaran *et al.*, 2012), an implicit assumption is made that the representation of mental health is given entirely by these responses. Analytically the researcher can go no further than to suggest that high scores on one measure are associated with high scores on another. If, however, the latent perspective is chosen, it is explicitly acknowledged that observed, manifest questionnaire responses are likely to be an imperfect realisation of an underpinning mental state. Given the central interest in the conceptual relationship between wellbeing and mental illness, and the existing assumption from Chapter 2 that each of the scales is unlikely to perfectly capture either of these constructs, this second interpretation is far more appropriate for this investigation.

To this end, and to demonstrate the empirical and substantive differences between the approaches, four separate and logically sequential analyses are carried out. Firstly, the simplistic, observational approach is taken – that is, the relationship between the observed Likert responses for each of the items on each of the questionnaires is modelled. This is done both for the aggregate score and the individual items, giving an overall correlation value and a correlation matrix giving the degree of relatedness between the raw responses to each of the items across both scales.

Secondly, it is important to know if the two scales actually capture the same thing, and therefore whether wellbeing should be considered as simply the inverse of mental illness. If they are being determined by the same underlying processes, then knowing the factor scores of individuals for either of the responses should enable prediction of their responses to the other. Consequently, and secondly in the sequence, the *latent* variables for the illness measure (GHQ-12) are regressed on the summed total *manifest* variable of the composite wellness measure (SWE). Characterising the relationship between the four, underpinning constructs of the GHQ-12 and the SWE will provide empirical evaluation of the similarity between the processes driving the illness measure, and the traditionally accepted (and supposedly unidimensional) understanding of wellbeing. In the evaluation of the relationship between the processes informing illness and wellness, the inverse must also be considered. For this the underpinning constructs of the SWE are regressed on the summed, manifest, total GHQ-12 scores, to analyse the predictive capacity of the underpinning processes driving wellness on the manifest realisation of illness.

This approach evaluates the relationship between wellness and illness as captured by the aggregate scale, but Chapter 2 has demonstrated the issues with simple aggregation to summed scores. Thus, the third step in the modelling sequence is to deconstruct the previous finding by separating the individual constituent items of each measure. This allows the investigation of the association between the underpinning processes of a given measure, and the manifest observations of each specific item. This involves regressing the four, latent constructs underpinning the GHQ-12 on the manifest observations of the individual items of the SWE. Doing so gives an indication of whether the underpinning processes of the GHQ-12 are associated with the outcomes of the SWE items. Using this it is possible to begin to assess whether the processes underpinning mental illness are the same as those driving mental wellbeing as captured by any individual item. For full confidence in addressing the relationship between the underpinning processes, the inverse must also be considered. Thus, additionally in the third step the latent underpinning constructs of the SWE are also taken and regressed on the manifest observations of the GHQ-12. This allows identification of whether there are any items in the GHQ-12 which seem to be associated with the same processes that drive the mental wellbeing response.

Finally, the fundamentally question in this chapter deals with whether the processes underpinning the two measures are related. As such, the fourth and concluding step must be to assess the relationships between the latent variables underpinning the wellness measure, and the latent variables underpinning the illness measure. This requires evaluation of the predictive capacity of wellness constructs for illness constructs, and vice versa. If the underpinning factor scores for each measure show little association, it may suggest that the underlying processes are indeed different for mental illness and mental wellbeing, and that researchers should treat them as separate constructs and not simply the inverse of each other.

3.2.2 Research Questions:

To characterise the SWE via ESEM, and then subsequently to evaluate the similarity between it and the GHQ-12 from each of the above methodological and substantive approaches, five research questions present themselves:

1. Can the Short Warwick Edinburgh Mental Well-Being Scale (SWEMWBS) be adequately understood using a traditional unidimensional interpretation?

2. If not, then what are the latent, unmeasured dimensions underpinning the SWEMWBS)?

3. To what extent are SWEMWBS (designed to capture mental wellbeing) and GHQ-12 (developed to screen for negative symptoms) correlated in observed outcomes? This gives rise to two questions:

    a. To what extent are the summed scores of SWEMWBS and GHQ-12 correlated?

    b. To what extent are the responses to individual items of the two scales correlated?

4. To what extent does this correlation in observed outcomes reflect "true" similarity between underlying latent constructs?

    a. What is the predictive capacity of the newly formulated underpinning dimensions of wellbeing as captured by the SWE on the manifest measurements of illness as captured by the GHQ-12?

        i. What level of predictive capacity do the underpinning, latent wellbeing dimensions of the SWE have for mental illness as captured by the traditional, unidimensional, summed GHQ-12?

        ii. What level of predictive capacity do the underpinning, latent wellbeing dimensions of the SWE have for mental illness as captured by the observed individual item responses of the GHQ-12?

    b. What is the predictive capacity of the four underpinning dimensions of mental illness as captured by the GHQ-12 on the manifest measurements of wellbeing as captured by the SWE?

        i. What level of predictive capacity do the underpinning, latent illness dimensions of the GHQ-12 have for mental wellbeing as captured by the traditional, unidimensional, summed SWE?

        ii. What level of predictive capacity do the underpinning, latent illness dimensions of the GHQ-12 have for mental wellbeing as captured by the observed individual item responses of the SWE?

5. Finally, and most importantly, to what extent can manifest wellbeing truly be considered the inverse of mental illness? How truly related are the underpinning processes of illness and wellness as captured by each measure?

        i. How correlated are the underlying constructs of the SWE wellness measure with the underlying constructs of the GHQ-12 illness measure?

3.3 Methodology

As in Chapter 2, Exploratory Structural Equation Modelling (ESEM) is carried out, a recent development in the Structural Equation Modelling Field, pioneered by Asparouhov and Muthen (2009). ESEM allows the relaxation of some of the more restrictive assumptions inherent in traditional SEM frameworks. This is particularly relevant in fields of psychological testing, where Exploratory Factor Analyses (EFA), Rasch Analyses (RA), Structural Equation Modelling (SEM) and Confirmatory Factor Analyses (CFA) are commonly used to evaluate the performance of psychological screening instruments (Marsh, 2007; Marsh *et al.*, 2009). As seen in Chapter 2, however, methodological advances and simulation studies have, in recent years, begun to cast doubt over the capacity of these traditional Item Response Models to critically evaluate proposed factor structures (Asparouhov and Muthen, 2009; Marsh *et al.*, 2014). Chapter 2 showed that this is particularly true for the more complex proposed factor structures, such as those multidimensional structures often suggested when evaluating psychological instruments (e.g. Worsley and Gribbin, 1977; Graetz, 1991; Marsh *et al.*, 2014; Rahmati Najarkolaei *et al.*, 2014).

In relaxing some of the more stringent assumptions of traditional Factor Analysis, ESEM allows the modelling of more realistically complex factor structures, allowing for more pertinent and realistic policy suggestions. Here the underpinning factor structure of the SWEMWBS measure is investigated, and subsequently evaluated against the previously identified and verified four-factor structure for the GHQ-12 from Chapter 2. Crucially, the capacity of ESEM to partition common and unique variance allows the investigation of the relatedness of the common underpinning constructs. As outlined above, this comparison of the SWEMWBS to the GHQ-12 is undertaken via four separate specifications; manifest on manifest, manifest on latent, latent on manifest, and latent on latent. This breadth and depth of specification gives the most complete possible picture of the total degree of relatedness between the two measures.

3.3.1 Model Specification and Estimation

The SWE responses are specified in the model as ordinal variables. As such the default estimator in Mplus is a robust weighted least squares estimator (Muthén & Muthén, 1998-2011). This requires a specification of seven probit regression equations, one for each of the constituent questions of the SWE (Muthén & Muthén, 1998-2011, pp. 62). These are interpreted as the log-odds of answering one response category higher (higher wellbeing) on the Likert response scale for that specific SWE item. This is then interpreted as measuring the strength of the relationship between the modelled underlying dimension and the probability of a change of one response category (Chen, 2007)[12].

As with Chapter 2, all models are run using both Maximum Likelihood and Bayesian estimation. This chapter then goes one methodological step further and regresses the individual estimated factor scores on both manifest observations and other latent unmeasured variables. For these results to be comparable and interpretable across these latent metrics, the results need to be standardised with respect to the underpinning latency. The subsequent units of the metrics are not empirically identical, however this standardisation still allows inference as to the unit change in outcome associated with a standard deviation change in latent metric.

As in Chapter 2, the outputs from the factor analysis are explicitly modelled as normally distributed scores with unit-variance. When interpreting the relationship between these standardised, latent variables and other both manifest and latent variables, this standardisation

[12] For a full discussion of rotation methods, distributional assumptions and Bayesian justification, see Chapter 2

influences the interpretation of resulting coefficients. The standardisation here refers to the explicitly normal latent variables which are modelled to have a mean of 0 and standard deviation of 1 by design. Exploiting this flexibility in latent specification allows results to be meaningfully compared across metrics. More specifically, where Latent unmeasured responses are regressed on manifest responses, coefficients are interpreted as the change in the latent response in standard deviation units caused by a unit change in the manifest response. Where latent variables are regressed upon each other in the latter section, the both dependent and independent variables are standardised. This means latent-on-latent coefficients are interpreted as the standard deviation change in response for a standard deviation change in predictor.

3.4 *Results*

3.4.1 Evaluating the need for multidimensionality

The first research question addresses whether the SWE can be adequately understood with a simple unidimensional interpretation. It thus needs to be established to what degree a more complex structure is necessary or useful. In order to evaluate this the unidimensional 1-factor solution is modelled, alongside 2- and 3- factor solutions and measures of relative badness of fit constructed for each. Table 3.2 presents the sequential improvement in model fit from adding further explanatory dimensions, starting with unidimensional interpretation and subsequently introducing explanatory factors. It is clear that across all measures of comparative badness of fit [13], and under both Maximum Likelihood and Bayesian estimation, the unidimensional specification, is worse than the two-factor interpretation, which is in turn worse than the three-factor solution. The 3-factor solution offers by far the best solution to the data. Furthermore, Yu and Muthén (2002) recommend an RMSEA cut-off of 0.05 for maximum likelihood estimated structures using categorical data. The 3-Factor solution is the only solution which meets this requirement.

Table 3.2 demonstrates very clearly that the traditional method of turning SWE responses into a single summed score does not do full justice to the constituent items on which it is based, there is clear interpretive and predictive capacity afforded by adopting an interpretation with a greater degree of complexity

[13] For a more comprehensive discussion see Chapter 2.

	Maximum Likelihood Estimation				*Bayesian Estimation*		
Model Structure	Chi Square	P-Value	CFI	RMSEA	Posterior P-Value	2.5% CI	97.5% CI
Unidimensional	14253.117	0.000	0.956	0.160	0.000	3539.202	3844.986
ESEM 2 Factor	2319.914	0.000	0.993	0.085	0.000	713.363	867.877
ESEM 3 Factor	264.356	0.000	0.999	0.047	0.000	151.170	236.826

Table 3.2: Model fit Statistics for Maximum Likelihood and Bayesian estimation of initial ESEM solutions for SWE data

3.4.2 The underpinning structure of the SWE

Having established that the 3 Factor structure fits the model significantly better than the simpler solutions, it is necessary to evaluate the construct validity by looking at the constituent factors. This is the motivation for the second research question. This was carried out by eliminating all loadings below 0.15, following the same procedure as for the GHQ-12 outlined in Chapter 2. The resultant factor structure with associated loadings, and factor correlations is given in Table 3.3[14]. Loadings are standardised probit regression coefficients, giving the predicted change in modelled latent construct with a unit increase in response category for a given item. Factor correlations are modelled correlations between the underpinning latent variables, between the standardised constructs, again each having unit-variance and mean-zero.

[14] Unit interpretation across the loadings and factor correlations are identical to Chapter 2.

SWEMWBS Items	F1 **Eudaimonia**	F2 **Ataraxia**	F3 **Focus**
SWE1 Optimistic	0.677	-	-
SWE2 Useful	0.849	-	-
SWE3 Relaxed	0.185	0.598	-
SWE4 Problems	-	0.897	-
SWE5 Clear Thinking	-	0.550	0.376
SWE6 Feeling Close	0.350	-	0.421
SWE7 Make Up Mind	-	-	0.864

Factor Correlations	F1	F2	F3
F1: **Eudaimonia**	1.000		
F2: **Ataraxia**	0.749	1.000	
F3: **Focus**	0.814	0.625	1.000

Table 3.3: Bayes estimates for the 3-Factor structure proposed for the interpretation of SWE responses

Estimated in Mplus using Bayesian estimation (100,000 iteration) and oblique Geomin rotation. Estimates given are standardised via a unit-variance constraint for the latent factors.

There are 3 separate and distinct factors identified by the analysis. The first factor has been termed Eudaimonia due to its similarities with the concept as considered in the literature. This factor deals with the individual's perception of their own purpose, associated with positivity via optimism, and worthwhile functioning via the highest loading of 0.849 for "Feeling Useful". It also loads on feeling close to others and feeling relaxed. "Feeling close to others" is one of the most interesting items as it is the only one that explicitly deals with individuals outside of the responding self.

Factor 2, termed "Ataraxia", meaning calmness or serenity, is associated with the ability to carry out tasks, without feeling overwhelmed – as evidenced by its highest loading on "Dealing with Problems Well" (0.897), and associated high loading on "Feeling Relaxed" (0.598). This suggests it is not just to do with the capacity to carry out tasks, but also to feel that those tasks are manageable. More simply, it is not solely whether tasks get finished, but also whether the individual feels they were within their capability.

The third factor is termed "Focus" and loads most strongly on the individuals' evaluation of their ability to make up their own mind (0.864). This initially simple notion of decision making clarity is interesting, as this item does not load on "Dealing with Problems Well", suggesting that there may be a nuanced difference between decision-making and the practical carrying out of these decisions, or that the decisions that individuals are making are not necessarily to do with problem solving. This might suggest that a high correlation between this factor and the social functioning elements of the GHQ-12 will be found. Focus also loads the most strongly on "feeling close to others" of the three modelled factors (0.421) suggesting some element of externality in the evaluation of an individual's capacity to make decisions and think clearly.

The modelled correlations between the factors are also given at the bottom of Table 3.3. The correlations are higher than the modelled GHQ-12 correlations seen in Chapter 2, but again the true amount of variance which would be predicted from one given the other is rather modest as this is given by the squared correlation coefficient. This means that even the high correlation between Eudaimonia and Focus of 0.814 means that knowing a complete set of factor scores from Focus would only allow you to correctly predict the scores of 66.4% of scores from Eudaimonia and vice versa. These factor correlations are below the thresholds commonly used in other SEM literature to assert unidimensionality (e.g. Gao *et al.*, 2004; Gouveia *et al.*, 2010; Fernandes and Vasconcelos-Raposo, 2013), so this more nuanced, multidimensional interpretation is carried forward into the next stage of evaluating the captured wellbeing scores against the GHQ-12 illness measure.

3.4.3 Investigating the Relationship between the GHQ-12 and the SWE

3.4.3.1 Manifest-on-Manifest Formulation

Having developed a more thorough understanding of the SWE measure, and what aspects of wellbeing it is capturing, it is necessary to now evaluate whether these elements can truly be thought of as the inverse of mental illness. It is worth reiterating at this point that the two measures are scored inversely. Higher scores on the GHQ-12 measure indicate higher levels of psychological *distress*, whereas higher scores on the SWE indicate higher levels of wellbeing. These interpretations are also true of the modelled factors underpinning the SWE and GHQ-12, with higher scores indicating greater wellbeing in the SWE constructs, and greater distress in the GHQ-12, with the notable exception of the GHQ-12 "Emotional Coping" construct as detailed in Chapter 2.

Firstly, the degree of similarity between the manifest elements must be evaluated. The correlation between the summed scores of the two measures in the data presented here is -0.607. This is almost identical to the -0.61 found by Mukuria et al. in 2014. Tennant et al. (2007) found in their initial development of the longer WEMWBS that the summed score correlation between it and the GHQ-12 was -0.53. It is perhaps unsurprising that the correlation here is stronger, given that the SWE is only half of the items of the full WEMWBS and the existing suggestion of their dissimilarity (Davies, 2014). This is a fairly strong correlation, and is, as expected, negative due to the scoring of the measures. The squared coefficient is 0.369, indicating that knowing the full summed GHQ-12 score for all individuals in the dataset would allow you to accurately predict 37% of the variation in SWE scores, thus whilst the correlation is fairly large, the predictive capacity afforded by the scales is rather low (Keyes, Dhingra and Simoes, 2010). This result is broadly as expected but suggests there is an empirical dissimilarity between the scales that it is of substantive interest to characterise.

It is now essential to look at the correlations between responses to individual items which are given in Table 3.4. Immediately noticeable is the low values for all items, despite the much more strongly correlated summed scores. Only 9 pairs of items have a true predictive capacity (given by the squared correlation coefficient) of over 10%, and none have over 20%. Given that simple correlation of the summed scores indicates a modest correlation (-0.607) it might be expected that there is a greater degree of similarity between individual items. The most strongly correlated items are SWE3 (Feeling Relaxed) with GHQ5 and GHQ9 (Feeling Unhappy/Depressed), but even these strongest correlations only represent true predictive capacities of 18.1% and 15.5% respectively. Very strikingly, the correlations between the questions that would be expected to be very strong, given the wording, are actually very low. For example – GHQ3 “Felt like you were playing a useful part in things”, and SWE2 “Feeling Useful” have a correlation of only -0.191. The possible causes for this are discussed further in the Discussion section below.

	GHQ1	GHQ2	GHQ3	GHQ4	GHQ5	GHQ6	GHQ7	GHQ8	GHQ9	GHQ10	GHQ11	GHQ12
SWE1	-0.097	-0.169	-0.126	-0.088	-0.151	-0.191	-0.127	-0.109	-0.248	-0.225	-0.195	-0.153
SWE2	-0.128	-0.203	-0.191	-0.116	-0.185	-0.227	-0.158	-0.131	-0.296	-0.314	-0.278	-0.165
SWE3	-0.163	**-0.357**	-0.117	-0.105	***-0.394***	-0.306	-0.204	-0.142	**-0.426**	**-0.362**	-0.277	-0.199
SWE4	-0.151	-0.299	-0.143	-0.131	-0.304	-0.314	-0.166	-0.171	**-0.387**	**-0.372**	-0.298	-0.187
SWE5	-0.175	-0.300	-0.134	-0.134	-0.303	-0.304	-0.163	-0.157	**-0.384**	**-0.380**	-0.303	-0.177
SWE6	-0.101	-0.187	-0.117	-0.083	-0.193	-0.200	-0.122	-0.104	-0.268	-0.254	-0.223	-0.147
SWE7	-0.123	-0.230	-0.112	-0.131	-0.230	-0.256	-0.120	-0.137	-0.291	**-0.316**	-0.257	-0.141

Table 3.4: Observed Correlations between Responses to Individual Items of GHQ-12 and SWE constructed from full 2009 wave of Understanding Society. Boldened values indicate pairs with <10% shared variance between items.

The inter-relatedness between the manifest responses to the questionnaires show some interesting characteristics. There are broad similarities, but the responses to individual questions differ markedly based on the questionnaire they are part of, even if the wording is almost identical. These differences are clearly not easily categorised. Furthermore, the difference between these similarly worded items demonstrates that the items are clearly not simply capturing the numerical inverse of each other. In order to evaluate this more completely the variance which is analysed here must be partitioned into common and unique variance via factor analysis. This takes us on to research question 3, the evaluation of the relationship between the commonalities underpinning each scale.

3.4.4 The Relationship between Latent Wellbeing and Manifest Illness

The following section considers the structure of each of the underpinning dimensions of each mental health questionnaire when explained using the manifest variable responses provided to the opposing questionnaire. Furthermore, it allows us to begin to assess the degree of similarity between the underpinning processes governing metrics capturing both "mental illness" and "mental wellbeing".

3.4.4.1 Latent wellness factors and summed manifest GHQ-12 responses

The first step in investigating the relationship between Latent Wellbeing and Manifest illness is to empirically evaluate the similarity between the identified wellbeing constructs from section 3.4.2 and summed GHQ-12 scores. This involves the creation of three correlation coefficients given in Table 3.5, reflecting the shared variance between the continuous, latent constructs of Eudaimonia, Ataraxia and Focus and total GHQ-12 scores. The overall degree of association is low. None of the absolute correlations are above 0.5, despite the correlation between the two summed scores being over 0.6. The correlations are uniformly negative, as expected given the scoring of the two metrics, higher scores indicate poor mental health in the GHQ-12 and good mental health in the SWE, this is reflected in the constituent constructs.

Latent Wellbeing on Manifest Illness	F1 Eudaimonia	F2 Ataraxia	F3 Focus
GHQ-12 Total	-0.372	-0.448	-0.347

Table 3.5: Correlation coefficients for modelled latent constructs underpinning the SWE and the manifest sum GHQ-12 score. Calculated from Understanding Society Wave 1.

Eudaimonia has a correlation of -0.372 with GHQ-12 scores. Squaring this gives a true predictive capacity of 0.138, implying that having a comprehensive list of estimated factor scores for Eudaimonia would only allow prediction of 13.8% of the variation in GHQ-12 scores. Individuals' evaluations of their own perceived self-realisation tell us very little about their experiencing of mental distress. This low correlation reinforces notions of the dissimilarity between the processes governing wellness and observed differences in mental illness.

Ataraxia similarly has a low correlation score of -0.448. Individuals who are more calm and relaxed as captured by the SWE do not necessarily have better mental health as captured by the GHQ-12. Again, squaring this correlation gives the true predictive capacity. Knowing full factor scores for Ataraxia would allow the accurate prediction of 20.1% of the variation in GHQ-12 total scores. This is again very low, although is the highest of the correlations here. This may be to do with the presence of a question in the GHQ-12 which specifically addresses feelings of relaxation, which Ataraxia reflects from the SWE.

The correlation between the SWE construct Focus and GHQ-12 scores is the smallest of the three. A correlation coefficient of -0.347 implies a predictive capacity of only 12%. Whilst being able to forecast an eighth of the variation in total GHQ-12 scores may seem important and informative in a purely mathematical sense, given the typical assumption of similarity between the processes governing mental illness and wellbeing, this is clearly very low.

3.4.4.2 Latent wellness factors and individual manifest GHQ-12 item responses

Whilst these correlations are uniformly very modest, as has been made explicit, Section 3.4.3.1 demonstrated the flaws with generalising to individual items from aggregate trends. The correlations in Table 3.5 cannot be assumed to imply uniformity in the GHQ-12 responses relationship with the wellbeing constructs. Indeed, such inference would be a variation on the aggregation fallacy, assuming uniformity within complex and additive processes. Thus, the next logical step in the evaluation of the relationship between latent wellness and manifest illness is to further deconstruct this by constituent SWE items.

Figure 3.2 illustrates regression coefficients between observed manifest responses, and standardised latent factor scores. Moreover, with results presented in Standard Deviation change on the latent scale for every response category change in observation, it is possible that larger coefficients will be found for the GHQ-12, which only has 4 response categories to distribute variability between, rather than the 5 of the SWE.

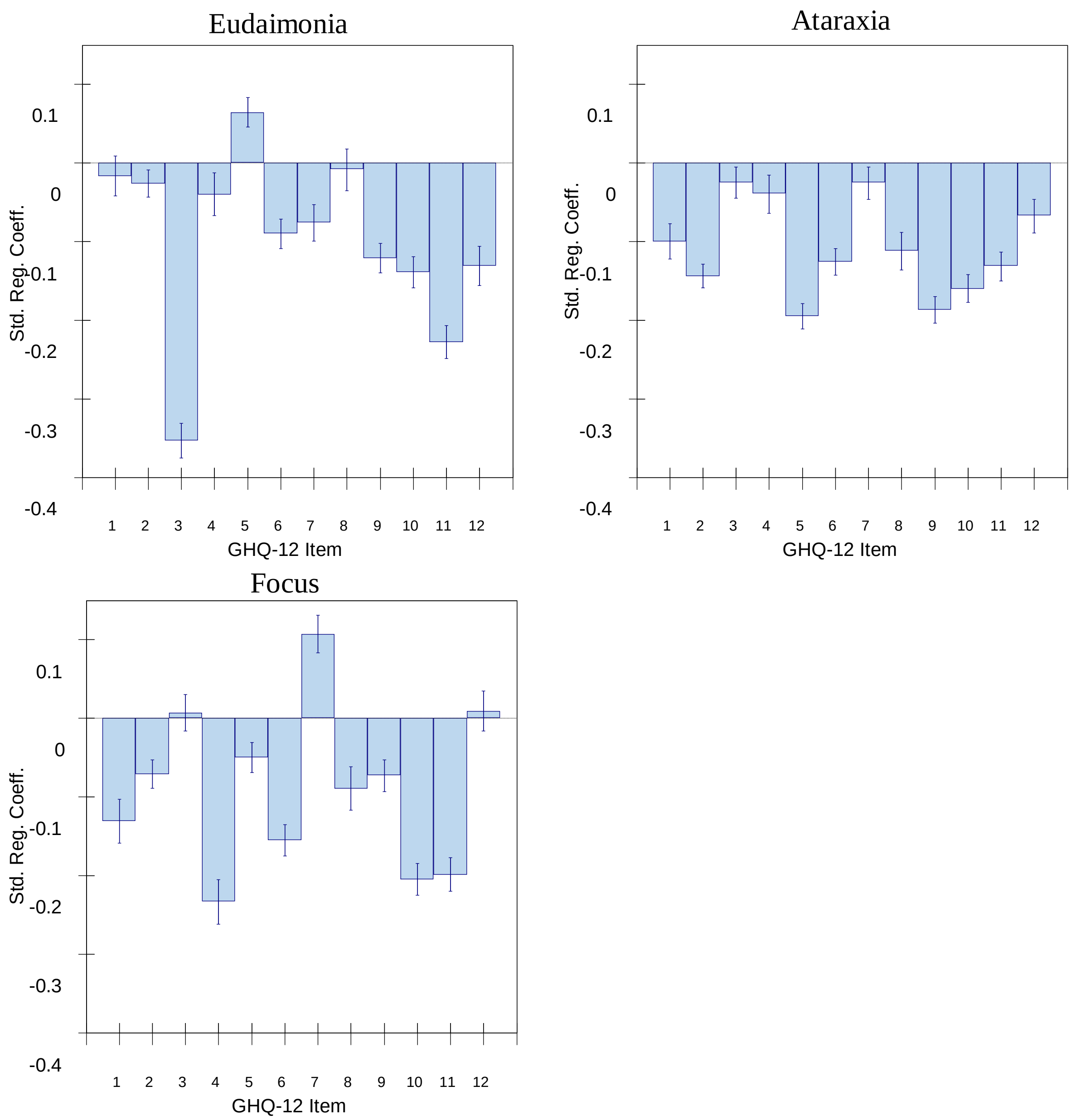


Figure 3.2: Graphs Displaying Standardised Regression Coefficients between each of the Latent SWE constructs and each of the Manifest GHQ-12 observations. The zero-value on the Y-axis implies no relationship, and if 95% confidence intervals span zero then this indicates there is no significant association at the 0.05 level.

The first construct, Eudaimonia, is shown to be most strongly inversely associated with GHQ Items 3 and 11. GHQ Item 3 refers to the individual's perception of whether they play a useful part in things. The strong relationship here is largely as expected, as the item that most strongly relates to this concept from the SWE questionnaire is a similarly worded Item, which asks if the individual is "Feeling Useful". The coefficient of -0.353 represents the change in standard deviation units for Eudaimonia for a unit change in GHQ3 response. Multiplying this across the four possible response categories, the total modelled variability possible here is 1.412, which corresponds to 51.98% of the variation in the normally distributed latent variable. This fairly strong relationship is also very consistent with the Eudaimonia concept that has been previously proposed in literature, as it refers to the individual's perception of their own sense of purpose. Items 9, 10 and 11 all refer to some element of lowered confidence or negative affect and have been commonly grouped under "lowered Confidence" in GHQ factor analysis literature (e.g. Graetz, 1991) so it is not necessarily surprising that where one of them is related to a construct the others tend to be also. The Eudaimonia construct is positively associated with GHQ5, which refers to feeling constantly under strain. This is interesting as it indicates those who score more highly on the Eudaimonia metric score more highly on feeling under strain, which tends to be considered as a negative effect. It is not inconceivable, however, that individuals with a greater sense of purpose also feel some degree of strain due to the expectation, values or pressure associated with this Eudaimonic sense of purpose. The Eudaimonia construct is unrelated to items 1 and 8, which refer to the individual's capacity to concentrate, and the capacity to face up to problems.

The second construct, Ataraxia, is not positively associated with any GHQ-12 item. It is most strongly negatively associated with Items 5 and 9, which refer to the individual's feeling of being constantly under strain, and their feeling of being unhappy or depressed. The Ataraxia construct is most strongly given by the SWE Item D which asks if the individual is "dealing with problems well?", with a loading of 0.896 as shown in Table 3.3. However, there is little association between the construct and the seemingly similar item in the GHQ, Item 8 – which asks, "have you recently been able to face up to your problems?". This stark contrast in the respondents' evaluation of questions reinforces the suggestion from Table 3.4 that there is more to the interpretation than simply the content of the item, such as response category wording, and framing of the entire questionnaire. The Ataraxia construct is also more strongly negatively associated with Items 2, 10 and 11. Item 2 refers to the individual's loss of sleep due to worry, which seems more consistent with what would be expected given the phrasing of the SWE Items that comprise the construct and their pertaining to relaxation and level-headedness.

The third SWE construct, termed "Focus", is most strongly negatively associated with Items 10 and 11 which refer to the individual's feeling of losing confidence and their perception of themselves as a worthless person. The Focus construct seems to be the most strongly correlated with the very negative, traditionally depressed elements of the GHQ-12. It is given most strongly by the SWE Item 7, referring to the individual's capacity to make up their mind, and is also the SWE construct that loads most strongly on "feeling close to others". The construct is also reasonably strongly negatively associated with SWE Items 4 and 6, which refer to the individual's capacity to make decisions as well as overcome difficulties, highlighting a link between individuals' perception of their own functioning and their personal confidence and security. The construct is noticeably positively associated with GHQ Item 7, which refers to the individual's capacity to "enjoy day-to-day activities". This positive association indicates that individuals who indicate that they enjoy day to day activities *less* in the GHQ-12, seem to indicate higher levels of Focus in the SWE. Further possible reasons for this relationship are speculated upon in the discussion. The Focus construct is also notably not associated with Items 3 and 12, which refer to the individual's feeling of usefulness, and their feeling reasonably happy.

There is clearly a complex relationship between the manifest GHQ-12 responses and the modelled latent constructs underpinning the SWE that have been developed in this Chapter. There are seemingly contradictory responses to similarly worded items, and the notion of positive experience as captured by SWE factors being associated with negative experience as captured by the GHQ-12 manifest observations seems to cast doubt over the simplistic, broad categorisation of these elements into positive and negative.

3.4.5 The Relationship between Latent Illness and Manifest Wellbeing

This section details results in the same format as the previous section. However, now manifest SWE responses are regressed on the standardised latent factor scores of the 4-Factor model of the GHQ-12 discovered in Chapter 2.

3.4.5.1 Latent mental illness factors and summed manifest SWE responses

As before, the first step is to empirically evaluate the similarity between the identified GHQ-12 illness constructs and summed SWE scores. Table 3.6 gives four correlation coefficients, reflecting the shared variance between the continuous, latent constructs of Lowered Self-Worth, Social Dysfunction, Stress and Emotional Coping and total SWE scores.

Latent Illness on Manifest Wellbeing	F1 Lowered Self Worth	F2 Social Dysfunction	F3 Stress	F4 Emotional Coping
SWE Total	-0.205	-0.221	-0.034	-0.106

Table 3.6: Correlation coefficients for modelled latent constructs underpinning the GHQ-12 and the manifest sum SWE score. Calculated from Understanding Society Wave 1.

The absolute correlations between the GHQ-12 and the summed total SWE scores are even lower than those for the summed GHQ-12 and SWE constructs. There is clearly little shared variance between the underpinning processes of the GHQ-12 and the summed wellbeing measure. These very modest correlations strongly suggest that the processes underpinning either construct are very different, or at least offer little predictive capacity over the other.

Lowered Self Worth is the first of the GHQ-12 constructs and has a correlation of -0.205 with total SWE scores. Experiencing low self-worth does not imply low wellbeing. The negative coefficient is a product of the expected relationship between the negative “lowered self-worth” construct, and the positively geared SWE. When squared this gives a true predictive capacity of 4.2%. This is very small, seeming to contradict the often-assumed similarity between mental illness and mental wellbeing.

Social Dysfunction has the largest correlation with total SWE score of -0.221. Experiencing low levels of social functioning does not imply low wellbeing. This again is very modest, offering a true predictive capacity of 4.9%. Knowing the full list of Social Dysfunction scores for the entire dataset of nearly 40,000 individuals would allow the accurate forecasting of the wellbeing of under 5% of individuals.

Stress has the smallest correlation, with only -0.034. The result was still significant however, the 95% confidence interval ranges from -0.010 to -0.059. The squared predictive capacity here is miniscule, having the full range of factor scores for stress would enable the accurate predicting of 0.12% of the variation in SWE scores. Knowing levels of wellbeing of individuals tells you close to nothing about their stress levels as captured by the GHQ-12. Whilst this correlation is very low, the generalisation cannot be logically made that SWE items individually have little or no relation to stress, as will become clear in the following section.

Emotional Coping again displays very low association with SWE scores. This suggests that stoicism as captured by the GHQ-12 is broadly unrelated to wellbeing as captured by the SWE. The correlation is once again very modest at -0.106. When squared this offers a true predictive capacity of 1.1%. Knowing Emotional Coping scores for individuals are tells you almost nothing about their wellbeing. This demonstrates the clear dissimilarity between the underpinning processes of the GHQ-12 with the observed SWE scores.

3.4.5.2 Latent mental illness factors and individual manifest SWE item responses

The next step is to consider individual items of the SWE and investigate their relationship with the GHQ-12 constructs. The results of this are provided in Figure 3.3

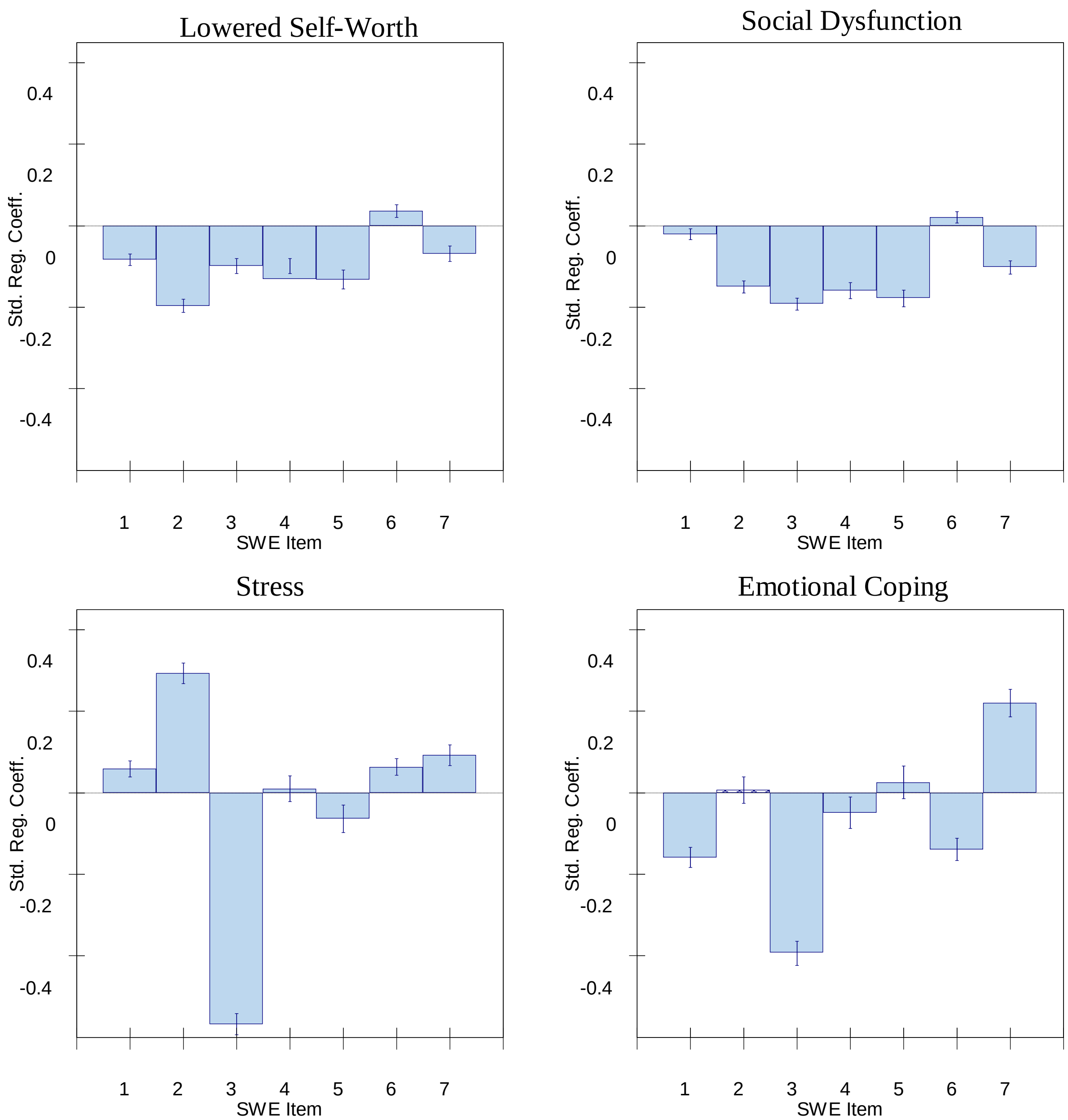


Figure 3.3: Graphs Displaying Standardised Regression Coefficients between each of the Latent GHQ-12 Constructs and each of the Manifest SWE Observations.
The zero-value on the Y-axis implies no relationship, and if 95% confidence intervals span zero then this indicates there is no significant association at the 0.05 level.

The Lowered Self-Worth construct is negatively associated with all but one of the SWE Items. It is most strongly negatively associated with Item 2, which is asks if the respondent is feeling useful. It is weakly, but significantly positively associated with Item 6, which asks if the respondent is feeling close to other people. The positive association suggests that those individual with lower self-worth are slightly more likely to be close to others, although this is clearly the least strong association. Lowered Self-Worth is also associated negatively and more strongly with Items 4 and 5, which refer to the individual's capacity to deal with problems well, and to think clearly respectively. This strong association with items that are referring to evaluation of performance would perhaps be expected to be more associated with Social Dysfunction and not self-worth, but this again illustrates the great deal of complexity and inter-relatedness within the constructs.

The Social Dysfunction construct is negatively associated with SWE Items 2, 3, 4 and 5. This is as expected as these items deal with pragmatic functioning, such as feeling useful, feeling relaxed, thinking clearly and dealing with problems well. The patterning is reasonably similar to that of Lowered Self-Worth. Relative to Lowered Self-Worth, Social Dysfunction is less strongly associated with Items 1 and 2 more strongly associated with Items 3,4 and 5. This indicates that relative to Lowered Self-Worth, Social Dysfunction is more associated with the items that constitute Ataraxia, and less associated with the items that constitute Eudaimonia, which would be expected to be expressed in the latent-on-latent section of results. Social Dysfunction is notably very weakly related to Items 1 and 6, which deal with optimism and feeling close to people, suggesting this is more clearly about social performance, than individuals evaluation of happiness.

The Stress construct is very strongly negatively associated with SWE item 3. This is related to feeling relaxed, which given the strength of the relationship, fits clearly with what would be expected from a Stress construct. Stress as captured by the GHQ-12 is also positively associated with Item 2 in the SWE, which deals with feeling useful. This seems to suggest that those who are stressed as captured by the GHQ-12, often feel more useful as captured by the SWE. This could suggest an interesting relationship where individuals who are more stressed tend to feel more useful, potentially as justification for the stress. Despite this association with feeling useful, the construct is notably not strongly related to dealing well with problems. Stress does not seem to be associated with an individual's evaluation of their own problem-solving capability.

The Emotional Coping construct is interesting for several reasons, the main reason being that it involves both negative and positive factor loadings, thus could serve to mask potential distress in summed scores. A more positive value indicates a tendency towards stoicism in the face of feeling strongly negative, which here can be framed as a negative outcome as it serves to mask distress. Emotional Coping is strongly negatively associated with SWE Item 3 – which again refers to "feeling relaxed". This leads us to the, perhaps intuitive, understanding that stoic individuals are less relaxed. It similarly implies that those who do not believe that feeling strongly negatively emotional affects their decision-making capacity (as defined in Chapter 2), tend to respond as feeling more relaxed. Similarly, the negative association with Item 1 suggests that more stoic individuals tend to be less optimistic. The positive association with Item 7 is also as expected given the characterisation of this GHQ-12 construct. High emotional coping was associated with individuals' feeling of being capable of decision-making even when feeling negatively emotional, so it is unsurprising that those with high scores on Item 7 – feeling "Able to make up my own mind about things" – are likely to have high levels of Emotional Coping. It is also interesting to note that the Emotional Coping factor is unrelated to Items 2 and 5 which refer to feeling useful and thinking clearly.

There is undoubtedly a complicated relationship between the observed manifest elements of either response and the underlying latent constructs underpinning the other. This is evidenced here clearly with the complexity of the relationships between emotional coping and SWE responses, although there are intuitive elements to these relationships, as there were between the manifest GHQ responses and latent SWE constructs. As outlined in the research questions, partitioning the variance in each response into that which is common across items and that which is unique to an item allows inference beyond the specific measurement with error, to try and access true, underpinning constructs. It is this relationship between the true, underpinning constructs of the GHQ-12 and SWE that is the next logical step in this sequence of analyses.

3.4.5.3 Characterising the relationship between the underpinning processes of Latent Wellness and Latent Illness

The final research question pertains to the relationship between the modelled latent factors underpinning each questionnaire. The factor correlations presented in Table 3.7 are interpreted much like the correlations from Chapter 2. However, here they are correlations between all 7 modelled dimensions underpinning the two separate questionnaires.

Whilst all correlation values here carry information, the values of greatest interest are in the lower left section – the relationship between the underpinning dimensions of the GHQ and the underpinning dimensions of the SWE. Again, it is worth bearing in mind that for the GHQ-12 factors, higher scores are associated with conditions typically thought of as “ill”. That is to say a higher score for “Stress” means you are more stressed. This is not true for Emotional Coping where higher scores indicate stoicism, which serves to insulate against certain people scoring more highly on the GHQ-12. The SWE constructs are all scored positively – with higher scores indicating greater wellbeing- which would be expected to produce negative correlations between the SWE factors and the GHQ-12 factors.

Factor Correlations		GHQ-12				SWEMWBS		
		Lowered Self-Worth	Social Dys-function	Stress	Emotion-al Coping	Eudai-monia	Ataraxia	Focus
GHQ-12	Lowered Self-Worth	1.00						
	Social Dysfunction	*0.687*	1.00					
	Stress	0.117	0.17	1.00				
	Emotional Coping	0.516	0.443	0.219	1.00			
SWE	Eudaimonia	*-0.529*	*-0.562*	0.049	-0.375	1.00		
	Ataraxia	*-0.550*	*-0.672*	-0.270	-0.416	**0.781**	1.00	
	Focus	-0.457	-0.571	-0.059	-0.169	*0.629*	**0.838**	1.00

Table 3.7Modelled Factor Correlations between Underpinning Dimensions of SWE and GHQ-12

Figures in italics reflect 25% of shared variance (|Corr|>0.5), and figures in bold reflect 50% shared variance (|Corr|>0.7). GHQ-12 factor correlations differ slightly from those in Chapter 2 as factor structure is evaluated here using data from only those who responded to the SWE (N=37,836)

Eudaimonia is clearly strongly and negatively correlated with the first two GHQ-12 factors, namely Social Dysfunction and Lowered-Self-Worth. The negativity is to be expected, simply as a product of the scoring. The correlations are of a similar magnitude, knowing a full list of the Eudaimonia factor scores from the individuals would allow you to predict just over a quarter of the variance in each of the dimensions. Substantively this suggests that individuals who are struggling to maintain social function, and who have very low self-worth, are likely to have low levels of Eudaimonia, that is – feeling like they have a purpose, and the things they do are worthwhile. Notably, the Eudaimonia construct is almost entirely unrelated to the stress construct, suggesting stress does not influence people's perception of their purpose. Finally, Eudaimonia is reasonably strongly negatively associated with Emotional Coping, suggesting that those individuals who consider themselves to have a worthwhile purpose are more likely to perceive negative emotion as affecting their decision-making capacity. Put simply, those who consider themselves worthwhile are more capable of acknowledging the negative impacts of feeling negatively emotional.

The Ataraxia construct is similarly strongly and negatively associated with the first two constructs from the GHQ-12. The correlation between Ataraxia and Social Dysfunction is the largest of the correlations between the two questionnaires at -0.672, meaning that knowing the factor scores for Ataraxia would allow you to predict just under 50% of the scores for Social Dysfunction. This fits with Figure 3.3 where the items constituting Ataraxia were more strongly related to the Social Dysfunction construct than Lowered Self-Worth. Lowered Self-Worth itself has a slightly weaker but still strong association with Ataraxia. This seems to suggest, very intuitively, that those individuals who think they are the least clear thinking and least capable of dealing with problems are those most likely to feel they cannot maintain optimal social function, which has a conceivable causal pathway either direction. Similarly, individuals with high levels of Ataraxia are unlikely to have issues with self-worth. The relationship between the Ataraxia construct as captured in the SWE measure, and the Stress construct as captured in the GHQ-12 measure is not particularly strong. Whilst there is a negative association between feeling calm and feeling stressed, this is not as strong as might be expected given the phrasing of the questions. Ataraxia is the SWE construct most strongly associated with Emotional Coping. This is likely to be indicative of the same mechanism that underpins the negative relationship between feeling relaxed and stoicism. The Ataraxia construct is also given by individuals' perception that they are dealing with problems well, so it is interesting that this is negatively associated with the Emotional Coping construct. This perhaps suggests that stoic individuals are in the minority. It is intuitive that feeling negatively emotional is strongly correlated with perception of impaired decision making across both measures.

Focus is the construct in the SWE least associated with the GHQ-12 outcomes. It has a strongly negative association with Social Dysfunction. This implies, again intuitively, that individuals who feel focused and capable of making up their own mind, are least likely to feel they cannot perform socially. Focus is less strongly associated with Lowered-Self Worth, although the relationship is still fairly strongly negative. The correlation of -0.457 means that knowing the information given by the Focus construct would only accurately allow prediction of 20.8% of the Self Worth of individuals. The Focus construct is again almost unrelated to Stress. The capacity of an individual to make up their mind seems unrelated to their evaluation of themselves as a stressed person. This seems consistent given the relatively large correlation between Focus and Social Dysfunction, and the lack of correlation between Social Dysfunction and Stress. Focus is also weakly related to Emotional Coping. This suggests that, perhaps unsurprisingly, individuals who consider themselves capable of making up their mind are slightly more likely to also be the stoic individuals who think that feeling negatively emotional does not affect their capacity to make decisions.

It is clear that whilst there is a reasonable amount of inter-relatedness between the mental health constructs modelled here, the overall predictive capacity afforded by each for the others is low. The largest correlation, between Ataraxia and Social Dysfunction, gives a true predictive capacity of less than 50%. This clear dissimilarity between underpinning processes has been echoed throughout this chapter. Overall there are substantively and empirically different processes going on in each of the measures, with evidenced differences between the constructs governing dimensions of illness and wellness.

3.5 Discussion:

This discussion speculates on some of the more interesting findings, and presents possible (but very tentative) suggestions as to what these results imply. However, this is done with an awareness that this is a cross-sectional study, and that the findings are only of associations between the mental health constructs discussed. Additionally, it is worth stating that the research defers to the psychological literature, and ultimately aim to only offer statistical and perhaps larger scale contextual insight to inform their claims.

The absolute item correlations illustrated that the most strongly correlated responses across the measures were the effect of feeling strongly emotionally negative on the individuals' feeling of relaxation, this has a conceivable causal pathway in either direction. The finding that there is a very weak correlation between SWE2 and GHQ3, which both ask a variant of "Are you feeling useful?", immediately highlights the need for greater understanding in the response patterning of the items. Given the very similar nature of the items, and the fact that they are only 6 pages apart in the self-completion questionnaire, with the GHQ-12 coming first (Understanding Society, 2011), the responses would perhaps be expected to be more similar. When considering the difference between these items, it is worth noting that SWE Item responses are phrased in terms of absolute rates of experience, that is "All the time" to "None of the time", GHQ-12 items are phrased in terms of relative burden, that is "Worst it has even been" to "Best it has ever been". This subtle shift in framing carries large differences in interpretation. Investigating the effects of wording differences is not a new avenue (e.g. Borgers, Hox and Sikkel, 2004; Dolan and Peasgood, 2006; Hankins, 2008; Smith *et al.*, 2013), although there is little written on the importance of the temporal implications of asking for relative as opposed to absolute consideration. It should be acknowledged that this frame-specific difference in interpretation may be part of the cause of the variation in responses.

We then investigated the relationships between the underpinning dimensions of each measure, with the observed scores for the other. This first involved the consideration of the relationship between the measures as aggregate scores with the underpinning latent scores of the other. Very modest correlations were found across all relationships. These relationships were notably weaker when considering the underpinning processes of the GHQ-12 and the observed summed score of the SWE. Factor correlations across the past two chapters have indicated that the GHQ-12 is more heterogeneous than the SWE, and these correlations reflect this. The SWE, with its fewer questions, seems more internally consistent.

The most interesting patterning in the Eudaimonia response was the positive association with “feeling under strain”. Whilst Eudaimonia can be considered something that is broadly positive, individuals tend to consider feeling under strain as negative. However, this seems to suggest that strain should be reframed as something not necessarily negative, or certainly not exclusively negative, and could potentially be the product of self-belief enabling individuals to undertake more, even if it is perceived as strenuous. It is entirely possible that strain can be associated with positive outcomes if the individual considers the strain worthwhile. This is further reinforced by the lack of association between Eudaimonia and the GHQ-12 Stress construct. Strain as captured by the GHQ-12 seems to represent solely negative experience of strain. Perhaps it is worth considering strain as having both positive and negative elements, pushing individuals to achieve more but ultimately becoming a burden if too intense.

There is a smaller than expected correlation between the “Ataraxia” construct – most strongly given by SWE Item 4, “dealing with problems well”, and the GHQ-12 Item 8 regarding “facing up to your problems”. The immediate suggestion might seem to be that these are due to wording effects. For example, “problems” in the SWE phrasing seem to reflect external problems that one encounters, where the possessive GHQ phrasing of “*your* problems” seems to reflect perceived internal issues. Thus, whilst these questions initially seem similar due to wording, the interpretation could actually be quite dissimilar.

The Focus construct is most strongly associated with the individual's perception of their capacity to make up their mind, however, it is also the construct most strongly associated with "feeling close to others". Whilst this is speculative, it seems to suggest that individuals' perception of their own competence is strongly informed by their feelings of security and support from others. This would suggest that elements of self-confidence, such as perceived competence, are potentially informed by external validation more commonly than they are by internal factors. It is also notably negatively related to enjoying day-to-day activities. This could suggest that individuals who do not enjoy their day-to-day activities still have high levels of focus and are buffered from this negative experience by being close to others. This might in turn suggest that the positive effect of being close to others is insulative against the apathy fostered by not enjoying day-to-day activities

The GHQ-12 Stress construct demonstrates particularly intriguing characteristics. It is predictably very strongly negatively associated with feeling relaxed, despite having almost no relation to the aggregate summed SWE. It is not strongly associated with the individual's perception of dealing with problems well, despite the strong association with "feeling useful". The relationship between Stress and perception of usefulness could suggest that individuals justify stress with the assumption that it, or their experience of it, must be useful. We would have perhaps then expected to see a relationship between stress and Eudaimonia, as this would solidify this suggestion, where individuals experiencing stress feel a stronger sense of purpose. This was not borne out, and in fact displayed the weakest association of any of the cross-measure correlations. The differences in the relationship between perceived stress and perceived usefulness for positive and negative mental health are clearly very complex and require further investigation. Again, this justifies thinking about reframing stress and strain as complex constructs which may not be uniformly negative.

Stress has notably low associations with any of the SWE constructs across the board. This suggests that the capturing of stress in positively and negatively geared questionnaires differs markedly. It is not possible from the responses to identify whether this is specifically due to differences in interpretation due to item phrasing, or whether these are truly different constructs. Again, we suggest that the perception of stress is not uniformly negative due to its association with some positive outcomes such as Eudaimonia. Further exploring what is inferred from “Stress” by respondents, in terms of relaxation or in terms of worthwhile exertion would be necessary to further understand the impact of stress on mental health.

Despite its negative loading for the GHQ-12, Emotional Coping behaves similarly to the Lowered Self Worth and Social Dysfunction constructs suggesting again that the negative emotional experience associated with the construct outweighs the stoicism associated with the negative loading for decision making. The lack of relationship between “thinking clearly” and Emotional Coping could potentially suggest that the negatively emotional “stoic” decision makers do not necessarily make judgements in a clear state of mind. They may even be aware of their inability to make rational decisions in times of negative emotions. If extrapolated this perhaps suggests of the Emotional Coping construct that those characterised as anxious worriers do not jump into decisions when they know they are negatively emotional, whereas the “doers” decide regardless. Individuals with high Emotional Coping values perhaps consider themselves capable of decision making regardless of quality of the decision.

3.6 Conclusion

Compared with the GHQ-12 structure and constructs discussed in Chapter 2, the SWE is more internally consistent, however the highest between factor correlation is still only 0.815, and the lowest is 0.618. Even the largest correlation offers a predictive capacity of only 66.4%, and the 0.618 correlation reflects a true predictive capacity of 38.1%. Assuming unidimensionality, and thus that the processes driving all items are identical, is akin to treating these as though they offer a perfect predictive capacity of 100% because it states that there is nothing to be gained from a more nuanced interpretation. Even in this case where there are only 7 constituent items in the questionnaire, with thematically similar content by design, these correlations are not high enough to justify this interpretation. There are clearly multiple dimensions to positive wellbeing as captured by the SWE.

Stratifying the SWE into three constituent dimensions gives us a greater chance of being able to identify real relationships with the underpinning processes of the GHQ-12. The raw correlation between the two summed scores, as they are typically used, is 0.607. We go further here, interrogating more comprehensively the degree of similarity between these two mental health questionnaires, and their assumed “illness” and “wellness” function.

We can see from the relatively low standardised regression coefficients across Figures 3.2 and 3.3 that the degree of predictive capacity afforded by understanding one of the dimensions underpinning either the GHQ-12 or the SWE, tells you little about the manifest observations of the other. Units are interpreted as the effective change in the manifest Item (in terms of standard deviations of the latent response), that you would expect to see with a unit change in item response. This means that the largest coefficient across either measure, which is SWE Item 3 (Relaxation) with the GHQ-12 dimension of Stress of -0.56, implies that a change in SWE Item 3 of one response category, would predict a change of just over half a Standard

Deviation, in the explicitly normally distributed, latent Stress construct. Across the five response categories of the SWE items, this means a total change of 2.8 Standard Deviations across the stress construct. Given the normality of the GHQ-Stress construct, we can infer that the total range of the SWE3 item can account for 83.85% of the total response variation in the "Stress", which is substantial predictive capacity across the measures. The strongest relationship between the SWE constructs and the GHQ measures is for the usefulness Item with the Eudaimonia construct. Whilst this accounts for considerably less variation than the aforementioned relationship, it still illustrates these are clearly related measures.

When considering the relationship between the latent variables, we see that there is evidence of some similarity between the first two dimensions of the GHQ-12 and the SWE, however again these are fairly small once squared to give true predictive capacity. We have speculated as to what these patterns may be indicative of, but we are truly asking here about the similarity between the measures, rather than seeking to explain the differences.

Considering the relationship at multiple conceptual scales highlights the heterogeneity in mental health experience and within summed aggregate measures. Particularly investigating the relationship between the underpinning dimensions of either measure and the manifest observed, summed scores of the other highlights some key methodological and substantive issues underpinning measurement research which are analogous to the aggregation fallacy. That is, averages or summations can be misleading when inferring to the compositional elements that constitute the average or summation. This highlights the potential pitfalls inherent in the prevailing wisdom of using aggregate measures. It furthermore highlights the logical fallacy of inferring anything about the patterning of constituent items constructing measures based on relationships expressed by the summed construct. Whilst this argument was made in the context of Chapter 2, where it became clear that within a single measure such as the GHQ-

12 there is considerable heterogeneity. It is clear that this becomes even more problematic when evaluating the relationship of that composite measure with other external variables. It is worth focusing in on a more explicit example:

Stress as captured in the GHQ-12 was almost entirely uncorrelated with wellbeing, as captured by the summed SWE. This was reflected in the associations between the SWE constructs, which failed to capture much of the variation in stress, as captured by the illness measure. It would be understandable but false to infer from this that the SWE fails to capture anything associated with Stress. However, this is immediately demonstrated to not be the case. When considering the relationship with individual items of the SWE the stress construct displays some clear and strong relationships. For example, Stress is very strongly negatively associated with feeling relaxed. However, this is not expressed in the correlation for the aggregate SWE score, as stress is positively correlated with the remaining SWE items. This calls into question the wisdom of using aggregate measures, or assuming that anything about the items constructing those measures can be inferred from the relationships expressed by the summed construct.

Overall in the relationship between the two measures individuals that indicate high wellbeing as captured by the SWE are less likely to score negatively on the GHQ – but the predictive capacity of the factor scores on each other are fairly low. The highest correlation between any of the constructs (Ataraxia and Social Dysfunction) only gives the capacity to predict barely 50% of the scores of the other. Beyond this, there are certain constructs which are clearly captured much more strongly by one of the measures. There is little relationship between Stress or Emotional Coping and the SWE constructs. Similarly, the SWE Focus construct is poorly captured by the GHQ-12 constructs. As in Chapter 2, there is clearly much going on here that is missed by analyses treating mental health as a broad, aggregate, monolithic construct.

Treating either questionnaire as unidimensional is clearly and demonstrably reductive. We further conclude, in line with a growing body of psychological literature, that assumptions of the similarities between the processes governing illness and wellness should at the very least be carefully scrutinised, if not altogether discarded.

4 CHARACTERISING THE DISTRIBUTION OF THE AGGREGATE QUESTIONNAIRES

4.1 What can be investigated in aggregate analyses of mental health?

Despite the increased awareness and acknowledged importance of mental health, there is still relatively little empirical research in the area, specifically for mood disorders such as anxiety and depression, particularly when contrasted with similarly burdening physical ailments (Tomlinson *et al.*, 2009). Where the previous chapters have been devoted to analysing what is being captured using conventional mental health metrics, this and the following chapter detail what can be modelled using those metrics, now that a greater understanding has been formed of what they are truly capturing. This chapter opens with a focussed overview of the existing findings of the extensive literature on at-risk groups for mental distress. This is then followed by a detailed outline of the research questions to be addressed and the methodology that has been applied in tackling them. While this chapter analyses the summed GHQ-12 and SWE cross-sectionally, the following chapter analyses the derived factor scores of the GHQ-12 longitudinally.

4.1.1 Individual Demographic Risk Groups

Previous research has shown there are several demographic factors which are known to be associated with mental health. As will become clear, demographic patterning often critically depends on the precise definition of illness used as a response. Furthermore, causality underpinning these relationships is far from agreed upon. Thus, a summary of some of the arguments concerning causality is also provided.

4.1.1.1 Gender Differences

Gender has been demonstrated to have a complex effect on mental health, with differing patterns for different mental conditions. Females have been repeatedly shown to experience worse mental health than males in traditional screening questionnaires (e.g. Weich & Lewis 1998; Propper et al. 2005). There is evidence however that this gender difference seems to hold more true for mood disorders than for psychoses (Häfner, 2003; McLean *et al.*, 2011), where gender differences are more commonly evidenced in age of onset rather than absolute cases (Jarema and Konieczyńska, 2001; World Health Organization, 2002; Memetovic, Ratner and Richardson, 2014). Furthermore, the pattern seems to be the inverse for death by suicide - with males suffering disproportionately (Uutela, 2010; Barr *et al.*, 2012; Reeves *et al.*, 2012; Reeves, McKee and Stuckler, 2014; Samaritans, 2015). Causal mechanisms for this have been proposed, such as stoicism in response to questionnaires (Murray *et al.*, 2008) and differential uptake of mental health services affecting the expression of mental distress in different outcomes (Pattyn, Verhaeghe and Bracke, 2015). There is little that can be done methodologically to separate out the effect of tendency to under-report, as identified by stoicism, from true lack of experience of negative symptoms. As such, it is hard to empirically characterise gender differences as either the expression of true differences in symptoms or differences in response tendency, in no small part due to the real explanation likely containing elements of both.

4.1.1.2 Age Differences

Age is also strongly associated with mental health, with a mid-life peak in mental distress being extensively identified and very widely accepted in mental illness literature (e.g. Propper *et al.*, 2005; Blanchflower and Oswald, 2008; Di Tella and MacCulloch, 2008; Clark and Oswald, 2011; Puustinen *et al.*, 2011)[15]. This has been suggested to be due to the accumulated stresses of increased or compound long term illness with age (Arokiasamy *et al.*, 2015; Cosco, Howse and Brayne, 2017), stigmatisation of ageing (Wurm and Benyamini, 2014) or increased workload and increased responsibility and stresses of parenthood (Paul & Moser, 2009).

Despite the broad acceptance and replication of finding this mid-life peak in distress, or U-shaped curve for wellbeing, it is not without question. The U-shaped finding is often asserted from cross-sectional studies (Blanchflower and Oswald, 2008; Chida and Steptoe, 2008; Wikman, Wardle and Steptoe, 2011) meaning the assumptions of longitudinal tracking along this U-shape are impossible to verify. Furthermore, the trend is typically identified using traditional population screening questionnaires (Propper *et al.*, 2005; Blanchflower and Oswald, 2008; Lang *et al.*, 2011). As was made clear in Chapter 3, translating these findings to wellness recommendations can be reductive, and the case has been made that this mental health peak is actually an artefact of a modelling process that fails to take account of cohort effects (Bell, 2014). Furthermore, the notion of a midlife peak in distress is notably contested in wellbeing literature (Steptoe, Deaton and Stone, 2015), with different dimensions of wellbeing being shown to demonstrate differing trajectories with age (Weich *et al.*, 2011). This is sometimes referred to as the "wellbeing paradox" whereby older individuals report higher wellbeing, despite experiencing conditions that seemingly predict negative mental health in other age groups (Windle, 2011, 2012; Cosco, Howse and Brayne, 2017).

[15] However whether this is truly representative of an effect of age is contested, see Bell and Jones (2015) for more information on the separation of age, period and cohort effects (Burton-jeangros and Editors, 2015).

4.1.1.3 Marital Status Differences

Married individuals have been suggested to be more mentally healthy than their unmarried or separated counterparts across numerous time periods and countries, with becoming married robustly associated with increased wellbeing and lowered psychological distress (Horwitz, White and Howell-white, 1996; Kim and McKenry, 2002; Frech and Williams, 2007; Williams, Frech and Carlson, 2010a). The driving mechanism for this is contested, but mostly considered to be a product of the satisfaction and support, and economies of scale that benefit long term cohabitation and marriage (Gove, Style and Hughes, 1990; Oppenheimer, 2000; Holt-Lunstad, Birmingham and Jones, 2008; Waite, 2009).

Again, this relationship is more complex than it initially seems, whilst wellbeing is significantly higher in those who get married and stay married (Horwitz, White and Howell-white, 1996; Keyes, 2002; Lindstrom and Rosvall, 2012), the effects are not consistent, and exit from marriage strongly predicts mental distress (Wade and Pevalin, 2004; Wilson and Oswald, 2005). Marital quality tends to deteriorate over time, and given that quality tends to predict the strength of the mental health benefit, it seems logical that the benefits of marriage also deteriorate with time (Williams and Umberson, 2004; Frech and Williams, 2007; Williams, Frech and Carlson, 2010b). Moreover, the theoretical models underpinning these relationships often cite stigma associated with marriage but the continual social informing of stigma means this relationship is far from temporally stable (Williams, Frech and Carlson, 2010b).

4.1.1.4 Employment Differences

The most easily categorised socio-economic proxy is unemployment. Unemployment has been found to be repeatedly and consistently predictive of poor mental health across numerous countries for the last 30 years (Ezzy, 1993; Paul and Moser, 2009; van der Noordt *et al.*, 2014). The finding has made headlines in recent years due to a large body of work demonstrating country level unemployment changes being strongly associated with increased suicide rates across a large range of countries in the wake of recent global recession periods (Stuckler *et al.*, 2009, 2011; Uutela, 2010; Kentikelenis *et al.*, 2011; Barr *et al.*, 2012; De Vogli, Marmot and Stuckler, 2012; Reeves *et al.*, 2012; Reeves, McKee and Stuckler, 2014). Beyond clinical outcomes the negative effect of unemployment has also been evidenced across both positive measures of wellbeing such as the SWE (e.g. Berkey, Frey and Stutzer, 2003; Di Tella and MacCulloch, 2008; Easterlin *et al.*, 2010) and for negative mental health outcomes such as the GHQ-12 (e.g. Ross, 2000; Propper *et al.*, 2005; Thomas, Benzeval and Stansfeld, 2005; Taylor, Pevalin and Todd, 2007).

Despite the ubiquity of the association, the causality of the relationship is often contested and the intricacies of the relationship are still not fully understood. Paul and Moser (2009) performed a meta-analysis on 237 cross-sectional studies of unemployment and mental health and demonstrated an overwhelming body of evidence to suggest that unemployed persons experience both more distressing symptoms and worse wellbeing than employed individuals. This has also been demonstrated to be true longitudinally, where employment transitions predict deterioration or improvement in mental health when leaving and joining the workforce respectively (Murphy and Athanasou, 1999; McKee-Ryan *et al.*, 2005; Flint, Bartley, *et al.*, 2013). The relationship is further complicated by the change in mental distress seeming to be moderated by the type of work the individual is leaving or joining and the differential stigma experienced by different demographics upon leaving employment (Artazcoz *et al.*, 2004; Paul and Moser, 2009).

Beyond officially registered unemployment, which is often criticised as an overly simplistic understanding of true labour market transitions (Flint, Shelton, *et al.*, 2013), individuals with lower income more broadly have been shown to have higher rates of mental health disorders (Lorant *et al.*, 2003, 2014; Mangalore, Knapp and Jenkins, 2007; Lang *et al.*, 2011). Furthermore, job insecurity has also been shown to have a strong association with psychological distress (Strandh, 2000; Ferrie, 2001; Flint, Shelton, *et al.*, 2013). Job classification is commonly used as a proxy for socio-economic status and has also been demonstrated to have a strong association with mental health with lower status workers (Ross, 2000; Muntaner *et al.*, 2004; Propper *et al.*, 2005). Socioeconomic status can quickly become conceptually woolly without definition in studies, so this chapter uses both unemployment as well as employment stratified by job classification to approximate socio-economic status.

4.1.1.5 Ethnic Differences

Ethnicity has been suggested to play a complex role in mental health with the effect differing according to measurement used. Studies focused on psychoses have commonly found that clinical outcomes such as involuntary detention (Swanson *et al.*, 2009; Vinkers *et al.*, 2010; Lawlor *et al.*, 2012; Singh *et al.*, 2012), care-seeking (Hunt *et al.*, 2015; McManus *et al.*, 2016) and substance abuse (McManus *et al.*, 2016) tend to find far worse rates of negative outcomes among non-white individuals than white individuals (Harrison *et al.*, 2009; Henderson, Evans-Lacko and Thornicroft, 2013; Davies, 2014). Common mood disorders, typically depressive, anxiety or panic disorders have also been shown to disproportionately burden non-whites (Jones-Webb and Snowden, 1993; Weich *et al.*, 2004; Chanfreau *et al.*, 2013; McManus *et al.*, 2016). Conversely, mood disorders as captured by population screening questionnaires often seem to contradict this patterning, with white individuals commonly being found to have very similar if not worse mental health than non-whites (Weich *et al.*, 2004; Chanfreau *et al.*, 2013;

Stewart-Brown, Samaraweera, Taggart, N. B. Kandala, *et al.*, 2015). Wellbeing literature largely concludes similarly, non-whites have tended to be found to have higher levels of depressive symptoms but not lower positive mental health (Hu *et al.*, 2007). Self-reported measures of wellbeing also seem to suggest those of black ethnicity commonly tend to report higher levels of mental wellbeing (Chanfreau *et al.*, 2013). Due to this complex, measurement-specific background there is little consensus as to the overall burden of mental illness by ethnicity.

The specific nature of the relationship is also contested. Ethnicity is commonly suggested in mental health literature to be acting as a broad but ultimately fairly poor proxy for deprivation (Williams *et al.*, 2010; Davies, 2014). The relationship between mental health and ethnicity in quantitative research is commonly found to be attenuated by the inclusion of other social characteristics, which contributes to this characterisation of the relationship as a product of the acknowledged effect of deprivation (Jones-Webb and Snowden, 1993; Jackson-Triche and Sullivan, 2000; McGuire and Miranda, 2008; Morgan *et al.*, 2009; Oguz, Merad and Snape, 2013). Furthermore, complex and ethnically-patterned social phenomena such as discrimination (Karlsen and Nazroo, 2002; Bhui *et al.*, 2005; Karlsen *et al.*, 2005; Berg *et al.*, 2011), mistrust (Whaley, 2001), isolation (Halpern, 1993), displacement (Porter and Haslam, 2005) and stigma (Corrigan *et al.*, 2012; Schomerus *et al.*, 2012; Assari, Watkins and Caldwell, 2015; Clement *et al.*, 2015) have all been shown to play a large role in psychosis and common mental disorders. Difference in exposure to these negative outcomes is often patterned by ethnicity, and thus ethnicity is often and plausibly viewed as a proxy for these negative characteristics (Jones-Webb and Snowden, 1993; Karlsen *et al.*, 2005; McGuire and Miranda, 2008; Singh *et al.*, 2014).

4.1.1.6 Educational differences

Higher levels of education have been suggested to be both insulative against poor mental health and predictive of positive mental health (Ross, Reynolds and Geis, 2000; Propper *et al.*, 2005; Subramanian, Kim and Kawachi, 2005; Chevalier and Feinstein, 2006; Bleil *et al.*, 2008; Barry, 2009). Low levels of education are also predictive of poor mental health (Grundy and Sloggett, 2003; Hu *et al.*, 2007). Furthermore, it has been suggested in the UK that the poorly educated are bearing the brunt of population-level increases in mental distress (Barr, Kinderman and Whitehead, 2015). The relationship between education and mental health is often characterised as non-linear. This non-linearity tends to take the form of diminishing returns, with additional educational qualifications being seen to have lessening benefits for mental health (Chevalier and Feinstein, 2006; Bracke, Pattyn and Von dem Knesebeck, 2013; Bracke, Van De Straat and Missinne, 2014).

As with previous variables, causality is also contested due to its association with socio-economic position and thus deprivation (Dolan and Peasgood, 2006; Katikireddi, Niedzwiedz and Popham, 2012). This leads similarly to the effect of education being commonly attenuated by the introduction of other demographic variables (Dolan and Peasgood, 2006; Mezuk, Myers and Kendler, 2013; Stewart-Brown, Samaraweera, Taggart, N. B. Kandala, *et al.*, 2015). The nature of the relationship is further complicated by the fact that mental health has been shown to be both cause and effect in poor educational attainment (Fergusson and Woodward, 2002). This can result in feedback effects in the expression of depressive symptoms and poor attainment (Gareth and Wilson, 2018). Furthermore, genetic studies have attempted to unpack the causality of the educational relationship and seem to favour social causation hypotheses over genetic determination, but there is little overall consensus (Mezuk, Myers and Kendler, 2013).

4.1.1.7 Housing Tenure Differences

Housing Tenure has been shown to also have a strong predictive capacity for mental health, individuals owning their own property have been consistently demonstrated to enjoy better mental health than renters (Hiscock *et al.*, 2003; Taylor, Pevalin and Todd, 2007; Meltzer *et al.*, 2012; Pierse *et al.*, 2016; Clapham, Foye and Christian, 2017; Hsu, Chang and Yip, 2017). However, the causal inference that can be made about housing tenure is limited for the same reasons as the previous socio-economic status, housing tenure is associated with a large number of socio-demographic characteristics (Gibson *et al.*, 2011; Mason *et al.*, 2013). Furthermore, there is no clear unidirectional causal relationship as changes in housing tenure can conceivably be modelled as both a cause and effect of poor mental health (Jelleyman and Spencer, 2008; Baker, Bentley and Mason, 2013; Morris *et al.*, 2017). This leads to difficulty in characterising the relationship between tenure and mental health as; truly causal, where lack of home-ownership causes poor mental health; or simply compositional, as a product of individuals already predisposed to poor mental health tend to live in rented accommodation (Baker, Bentley and Mason, 2013).

Inference is also hampered by the social constructs informing the possible effect of housing tenure. As with previous variables, housing tenure may be important simply because it acts as a marker for income, security, isolation and self-esteem (Hiscock *et al.*, 2003; Bentley *et al.*, 2011; Suglia, Duarte and Sandel, 2011). Furthermore these socially constructed expectations suggested to underpin the relationship with housing tenure have themselves been demonstrated to be patterned by demographic characteristics (Tomaszewski and Perales, 2014) Furthermore, some studies have actually found the opposite relationship to the one described above. For instance, when becoming a homeowner is associated with unaffordable mortgage payments, then home-ownership is actually associated with a worsening of mental health (Taylor, Pevalin and Todd, 2007; Bentley *et al.*, 2011). This seems to lend strength to the suggestion that the relationship between mental health and tenure reflects a relationship between perceived security and mental health. As with other demographic variables, the complex causality underpinning the relationship between mental health and housing tenure is of demonstrable importance because different explanations carry very different implications for policy (Clapham, Foye and Christian, 2017).

4.1.2 Complexity in Characterising Demographic Risk

Whilst there is some clear consistency in the findings of these bodies of literature, it is clearer that there is still far more to be explored in understanding the demographic patterning of mental health. Furthermore, there is evidence of different patterning for positive and negative mental health, echoing the findings of Chapter 3. These clearly demonstrate the need for further interrogation of these patterns using the most advanced methods available. Whilst demographic predictors have received much of the focus in studies of mental health determinants, usually at the individual-level, this is by far from the only avenue of prediction that has been explored.

4.1.3 Contextual Geographical Risk

There is a long-standing literature on the ecology of mental illness, seeing where people live as of importance in determining their mental state. This dates back at least to 1939 (Faris and Dunham, 1939), and the ecology of mental illness remains the focus of many investigations to date (Mair, Diez Roux and Galea, 2008). Moreover, previous studies have demonstrated that factors associated with mental illness can occur at more than one spatial scale (Ross, 2000), with systematic reviews confirming the presence of neighbourhood contextual effects in mental health outcomes (Mair, Diez Roux and Galea, 2008). However, despite recommendations that it should move in a similar direction (Ryan and Deci, 2001), wellbeing research has not seen the same level of spatial investigation.

4.1.4 Complexity in Contextual Effects

People are located in complex environments, and it is of interest to determine and disentangle the effect of each of these. These multiple environments can be conceived as operating at multiple scales – from that of the household; to that of the neighbourhood; to that of the broad regions in which they live. Consequently it seems essential that analysis of mental health processes should not be constrained to a single spatial scale (Pickett and Pearl, 2001). It is important to know at which spatial scale to target policy-intervention. Given this potentially complex role of geographical context, it is important to specify a methodology able to identify which spatial scale is the most important in the predicting of mental health.

Given that geography has been demonstrated to play a role in mental health literature, another question presents itself. What is it about this geography that drives this relationship? It is of further interest to researchers to know whether differences in mental health are to do with the specific location of the mentally ill, or if there is some broader trend linking places over and above their specific location. Ultimately this requires distinguishing whether it is the location that is the important factor in geographical determination of mental health, or some attribute of the location. For example, notions such as availability of green space and associated urbanity, which have both been suggested to impact mental health (Sugiyama *et al.*, 2008; van den Berg *et al.*, 2010), can be grouped together and considered as part of the type of place itself. It is important then that the methodology for this analysis has the capacity not only to investigate the relative importance of geography but be able to ask questions of the nature of the relationship underpinning this importance.

The previous chapters have outlined the shortcomings of analytical simplicity in interpreting the GHQ-12 and the SWE. Given the twin methodological focuses of this thesis on measurement and modelling, this chapter explores the use of novel and advanced modelling techniques on the summed scores of the two metrics to illustrate what *can* be identified in traditional mental health studies. This shows the benefits of a suitably advanced methodology even with a potentially overly simplified response variable. Therefore, in order to demonstrate the benefit of adopting complex methodological specification regardless of measurement instrument, this chapter explores what can be found using the metrics as intended, in their unidimensional, summed forms.

4.1.5 Research Questions

Given the literature background, and the advocation of nuance in interpretation even with the mental health measures being used as they were intended, as summed unidimensional scores of illness and wellbeing, several questions present themselves:

1. How correlated are the two measures GHQ-12 and SWE at each geographical scale?
2. To what degree are the two responses geographically patterned, which scale is most important for each response?
3. How are each of the measures patterned with respect to a range of demographic predictor variables?
4. Having accounted for individual characteristics, does geography play the same role in mental illness and mental wellness?
    a. Which geographical scale is of greatest importance for each response?
    b. Is the strict location more important than the *type* of location in determining the responses?
5. Is it possible to identify specific locations and/or types of location that can be identified as having elevated risk of poor mental health?
6. Does geography matter equally for all?
7. Are these demographic and geographical patterns different for the two responses? Does the role of geography remain the same even after taking account of demographic characteristics?

4.1.6 Chapter Outline

The rest of this Chapter is organised as follows. Details of the data used are given and the modelling approach are first discussed - detailing the different specifications used for each research question. The chapter then goes on to detail the findings of the research, in the form of graphically illustrated model predictions structured by individual research questions before discussing the implications of these findings within a wider mental health context. Finally, some limitations with the current work are provided and suggestions are made for directing future research to overcome these.

4.2 Data

4.2.1 Understanding Society

This study analyses data from the first-wave (2009-10) of the Understanding Society (US)[16] dataset. Overhauling and incorporating the British Household Panel Survey (BHPS), US is the largest nationally representative survey available in the UK. Postcodes for primary sampling units (PSUs) were selected via multistage stratified selection (sorted on job classification and subsequently ethnic minority density) in England, Scotland and Wales. Addresses were subsequently selected using systematic random sampling from these PSUs, where addresses in Northern Ireland were selected at random without demographic stratification (McFall, 2011). This explicitly clustered design is important as it is exploited in the modelling process, as will become clear. The resulting dataset totals almost 51,000 individuals from across the UK.

4.2.1.1 Response Variables

The BHPS has been favoured in UK mental health studies due to the inclusion of the Likert 12-Item GHQ alongside detailed demographic and geographical information (Propper *et al.*, 2005; Jones and Wildman, 2008). Using the US dataset offers great statistical power alongside this mental health data, with 39700 individuals answering the GHQ-12 in 2009. This represented the largest distribution of mental health questionnaires in the UK at the time of distribution.

Furthermore, in the first wave of US, the SWE was also included. The majority of individuals who undertook the GHQ-12 also undertook the SWE (37836 of 39700) meaning that

[16] Also referred to as the UK Household Longitudinal Survey (UKHLS)

comparisons between the two can be readily drawn. Unlike GHQ, higher SWE scores in their raw state indicate greater health, so a simple arithmetic transformation was carried out to make them more directly comparable.[17] The distribution of this recoded variable is illustrated below in Figure 1 – showing a remarkably similar distribution to the GHQ-12 scores.

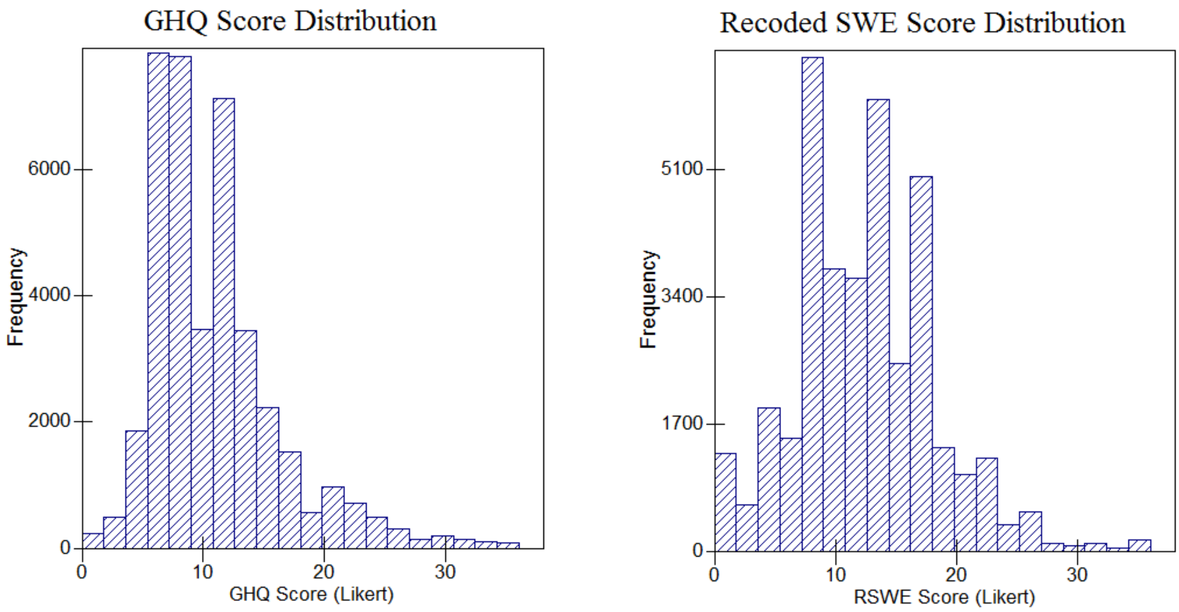


Figure 4.1: Histograms of GHQ-12 responses (left) and recoded SWE responses (right).

Several individual and household characteristics suggested from a comprehensive reading of previous work are included as variables of interest in the analysis (Table 4.1). Furthermore in order to analyse the effect of type of place over and above specific location, a measure of census tract classification is used to categorise areas into 52 different types of place based on socio-demographic characteristics (see Vickers & Rees 2006 for full description). Finally, 22 higher-level regions were constructed using a formulation from Jones et al. (1992); these are a

[17] GHQ-12 scores are calculated as the sum of responses to the 12 questions, each answered on a 0-3 scale with *higher* scores indicating mental distress, giving a total range of 0-36. SWE scores are calculated identically, but with 7 questions each scoring 1-4 with *lower* scores indicating mental distress, giving a total range of 7-28. They were recoded as follows to give numerically comparable scores:

$$Recoded\ SWE = \left((-1 * SWE) + 35\right) * \left(\frac{36}{28}\right)$$

breakdown of standard regions into highly urban "Metropolitan" and more rural "Non-Metropolitan" areas. These were then cross referenced with the PSU data to classify each of the PSUs into one of these higher-level regions.

INDIVIDUAL LEVEL DATA	**39700 Individuals**	
Age	Continuous; Mean = 45.81, Standard Deviation = 17.97	
Sex	*MALE* 43.9%	Female 56.1%
Marital Status	*SINGLE* 31.5%	Married 51.0%
	Separated/Widowed/Divorced (SWD) 17.5%	
Ethnicity (Ethnic)	*WHITE*: 83.4%	Black: 4.4%
	Asian: 8%	Mixed: 3.3%
Job Classification	*UNEMPLOYED* 43.5%	Professional 3.6%
	Managerial/Technical 21.0%	Skilled Non-Manual 12.7%
	Skilled Manual 9.0%	Partly Skilled 8.2%
	Unskilled 1.9%	
Highest Educational Qualification	*HIGHER EDUCATION* 34.0%	A-Level 19.3%
	GCSE 20.9%	Other Qualification 4.9%
	None 20.9%	
HOUSEHOLD LEVEL DATA	25181 Households	
Housing Tenure	*OWNED OUTRIGHT* 28.6%	Mortgaged 38.1%
	Rented 31.1%	Rent-Free 1.3%
	Other 1.0%	

Table 4.1 Descriptive statistics of variables used in the analysis and their subcategories within the dataset

Format: Categories, Percentage Coverage, [Observations]. For non-continuous variables those categories that are used as the base reference category are capitalised. Distributional statistics are given for continuous variables. Source: Calculations done by the author using data made available from Understanding Society Wave 1.

4.3 Methods

4.3.1 The Multilevel Approach

The data can be conceived as having a complex, multilevel structure as shown schematically in Figure 4.2. At the base of the structure are the two mental health measurements for each individual. These individuals are situated within their relevant household, local neighbourhood (defined by PSU) and broad geographical region. Moreover, each household is conceptualised as being part of a broader non-geographical construct of Area-Type. A multilevel model is needed for both technical reasons (to obtain valid standard errors) and for substantive reasons as the research questions require investigation at multiple levels. The underpinning model is ultimately a 5-Level Bivariate, Cross-Classified model. Predictor coefficients are interpreted as the difference in each response (GHQ-12 or SWE) associated with a unit increase in the predictor variable. Categorical variables are specified as a series of dummies, with the unit increase denoting the differential effect of belonging to that category relative to a reference category specified in Table 4.2.

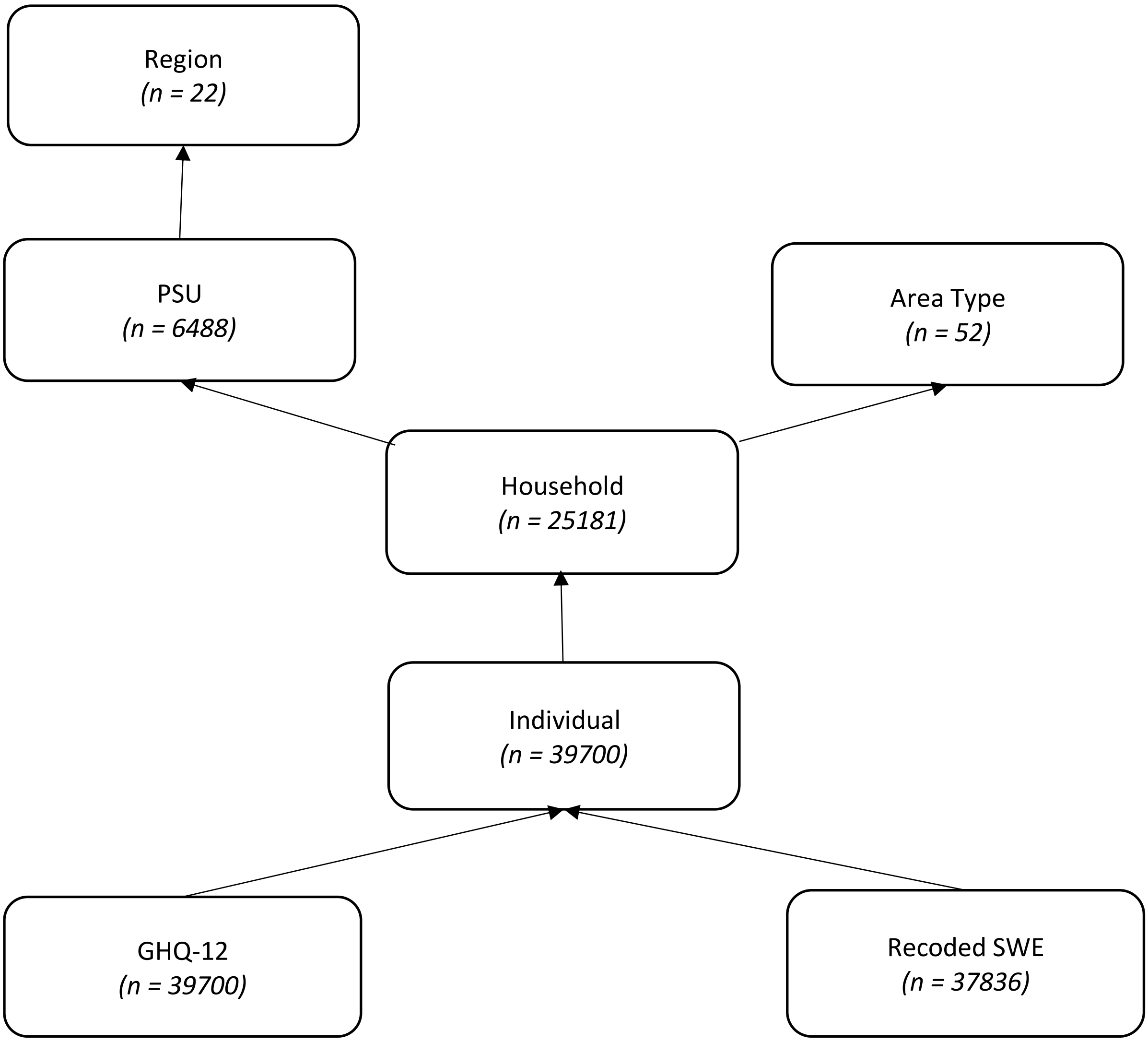


Figure 4.2: Illustration of Multilevel Structure of the Data

4.3.2 The Need for Multilevel Modelling

Multilevel modelling offers several useful methodological contributions here. Firstly, due to multilevel techniques capability for estimating variability between individuals and places simultaneously it is possible partition variance for all geographical scales, and thus investigate the relative importance of each (Jones, 1991; Hox, 2002). Specifying this complex geography also guards against overstating the importance of any specific spatial scales due to omission of relevant levels (Tranmer and Steel, 2001; Moerbeek, 2004). This is formalised via the specification of the Variance Partitioning Coefficient (VPC), which gives an explicit measure of the proportion of total variation in each response at each level, net of other structural levels.

Secondly, the multilevel model deals with the nature of the US survey which is clustered by design, with postcode sampling used to generate eligible households for the sample. This results in a potential dependence between observations based on geographical proximity. Individuals who are clustered within the same household or same PSU are more likely to respond similarly to each other with respect to each of the responses. This lack of independence would be problematic in a traditional regression analysis, which works under the assumption that all observations are truly independent (Goldstein, 1986). Ignoring this dependence risks overstating the real impact of predictors due to differential effects being measures of individuals within the same household especially in relation to the standard errors of the estimated coefficients. Multilevel modelling explicitly addresses this concern, simultaneously modelling at all structural levels and allowing for this potential dependency, thereby avoiding Type 1 errors and over-elaborate model specifications (Jones, 1991).

Thirdly, this analysis is concerned with a complex structure. Multilevel modelling typically analyses strict hierarchy in which the lower level units are conceived as nesting in one and one only higher level unit (Goldstein, Browne and Rasbash, 2002). Here however the structure is complicated by the need for the simultaneous analysis of households as part of a strict geographical structure, as well as belonging to a non-geographical grouping of types of place. Multilevel techniques can however be extended to such a cross-classified structure (Rasbash and Goldstein, 1994) to investigate the relative importance of functional and locational geography. This allows conceptual improvement from discrete understandings of contexts, to the consideration of a more realistic situation of multiple, overlapping contexts acting simultaneously (Arcaya *et al.*, 2012). Addressing locational and functional geography simultaneously, via area classifications as well as spatial nesting, allows this further partitioning of variance. This, in turn, allows quantification of the differences associated with the type of place individuals live in, rather than the specific place itself (Jones, Gould and Watt, 1998).

4.3.3 Basic Bivariate Model Specification

A model is required that estimates both responses simultaneously to facilitate comparison between the two outcomes at each scale. A simplified 2-level example of this multilevel bivariate structure is given below[18]:

$$Y_{1jk} = \beta_1 + (v_{1k} + u_{1jk})$$

$$Y_{2jk} = \beta_2 + (v_{2k} + u_{2jk})$$

$$\begin{bmatrix} v_{1k} \\ v_{2k} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma_{v1}^2 & \\ \sigma_{v1v2} & \sigma_{v2}^2 \end{matrix}\right)$$

$$\begin{bmatrix} u_{1jk} \\ u_{2jk} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma_{u1}^2 & \\ \sigma_{u1u2} & \sigma_{u2}^2 \end{matrix}\right)$$

Where Y_{1jk} gives the observed score for GHQ-12 of person j in household k, and Y_{2jk} gives the observed SWE score of person j in household k. The parameters β_1 and β_2 give the average score across individuals in a typical household for GHQ-12 and SWE respectively. Household differentials for GHQ-12 and SWE scores are given by v_{1k} and v_{2k} , and are assumed to come from a joint Normal distribution. The variance of the Level-2 residuals for GHQ-12 is given

[18] Note there is no subscript 'i' for the lowest level as this level exists solely to define the multivariate structure, and simply denotes whether a given response is for the GHQ-12 or the SWE.

by σ_{v1}^2, this represents the variation in GHQ-12 responses between households. The variance of the Level-1 residuals for GHQ-12 is given by σ_{u1}^2, which in turn represents the variation between Level-1 individuals within the higher-level units, these having already been accounted for by the Level-2 term. From this the Variance Partitioning Coefficient (VPC) can be calculated which assesses the proportion of total variation in the response accounted for by a specific level. This is calculated by the variance at a certain level given as a proportion of total variance summed across all levels. In this model the Level-2 VPC for the household level variation as a proportion of total variation across household and individuals is given by:

$$\text{Level 2 VPC for } Y_{1jk} = \frac{\sigma_{v1}^2}{\sigma_{u1}^2 + \sigma_{v1}^2}$$

Covariance terms assess how the residuals at each scale covary with respect to one another. Thus, the Level-2 covariance between the two scales is given by σ_{v1v2}, measuring the relationship between the Level-2 residuals, the household effects, around their respective means. Here it is assessing the extent to which households are high on illness are also high on well wellness. For instance, a large positive coefficient at Level-2 would imply that households which score highly for one measure are likely to score highly for the other. The Level-1 covariance between the two scales is given by σ_{u1u2}, measures the relationship between the individual level differentials for the two responses having taken account of the household effects. A high covariance implies that individuals who respond highly to the GHQ-12 would tend to respond highly to the SWE, irrespective of their household values. Standardising these values gives us the correlation coefficient between the residuals for each scale, bounded by -1 and 1. These standardised values are obtained by dividing the square root of the product of the associated variances. This allows inference about the similarity of patterning of the wellbeing and illness measure at the structural level of interest.

4.3.4 Extending the Basic Model

In order to answer the research questions several extensions to the basic model are required. A major extension is to include individual and household variables in the model.

$$Y_{1jk} = \beta_1 + \sum_1^m \beta_l x_{ljk} + \sum_1^m \beta_l x_{lk} + (v_{1k} + u_{1jk})$$

$$Y_{2jk} = \beta_2 + \sum_1^m \beta_l x_{ljk} + \sum_1^m \beta_l x_{lk} + (v_{2k} + u_{2jk})$$

$$\begin{bmatrix} v_{1k} \\ v_{2k} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma_{v1}^2 & \\ \sigma_{v1v2} & \sigma_{v2}^2 \end{matrix}\right)$$

$$\begin{bmatrix} u_{1jk} \\ u_{2jk} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma_{u1}^2 & \\ \sigma_{u1u2} & \sigma_{u2}^2 \end{matrix}\right)$$

Explanatory variables (the x's) can be included across both responses and can be included at the individual and household-level. The key terms in this so-called fixed part of the model are the β_l terms, which like standard regression coefficients measure the impact of a unit change in the explanatory variable on each of the responses. Unlike standard regression coefficients their standard error is corrected for potential dependence due to data structure. Their inclusion then alters the meaning of the random-part variances as they now assess the degree of unexplained variation at each structural level having taken account of these 'compositional' variables. The explanatory variables can, again, be categorical or continuous.

The data has several more structural levels, one of which is cross-classified – with households belonging simultaneously to PSUs and Area-Classifications. This is achieved by specifying further variances and covariances at each level and this allows the assessment of the size of effects net of these other scales for both outcomes. The estimated variances and covariances are necessarily sensitive to inclusion of demographic and structural variables. Thus, it is possible to address how the scales and similarities between the scales vary across structural levels as variances and covariances are estimated at every level in the model conditional on the predictor variables that have been included.

The analysis is concerned with identifying specific at-risk areas or area-types. To detect this, after the inclusion of demographic controls, residual differentials are examined at both region and area-type levels. This allows visualisation the estimated mental health of an area or type of area, relative to the nationwide average. Moreover, it is possible to plot the residuals of each response against one another to further analyse if the measurements are consensual – a positive gradient meaning that a higher score in one response implies a higher score in the other.

Finally, to determine the age-specific importance of geography, the overall fixed effect for age is replaced with an allowed-to-vary parameter. This allows the estimation of a variance term for age at different structural levels. This term is allowed to vary at both region and area-type levels, in order to ascertain which of the two had greater impact for each of the responses.

To answer the research questions the following 3 models were specified:

Model 1 – Fully specified 5-Level Cross Classified Model, no predictor variables. A null or empty model that simply partitions the variance around the two grand means of GHQ-12 and SWE overall scores.

Model 2 – Fully specified 5-Level Cross Classified Model, with additional fixed predictor variables. This is used to ascertain the mean effect of sociodemographic variables and explore what happens to the unexplained variances at the various levels when these compositional variables are taken in to account

Model 3 – Fully specified 5-Level Cross Classified Model, with a complex Age-term allowed to vary at Region and Type-of-Area levels (structural levels 4 and 5).

4.3.5 Model Estimation

As all models listed are cross-classified, they are estimated using Markov Chain Monte-Carlo (MCMC) simulation-based estimation [19]. This is because traditional likelihood-based estimation methods exploit the hierarchical nature of the data to achieve convergence and perform poorly in more complex situations (Browne and Draper, 2006). Models were run for 100,000 iterations, with a discarded burn-in of 5,000 iterations. Deviance Information Criterion (DIC) was used to gauge the effect of the inclusion of each model change, which measures the fit of the model whilst penalising for the number of included parameters (Spiegelhalter *et al.*, 2002) which are estimated as part of the MCMC procedure.

[19] Model assumptions expressed in previous section are given in terms of likelihood procedures, as these are how the starting values were obtained. In Bayesian methods, there is still the assumption of normality in the residuals, however the distribution of the variances and covariances are allowed to follow the more flexible and potentially skewed inverse-Wishart distribution.

4.4 Results

4.4.1 Relatedness of the two responses at different structural scales

The first research question addresses how consistent the two mental health responses are, and to what extent they are evaluating the same underlying dimension of mental health. The similarities between the scales are given by the correlation coefficients in Table 4.2. This extends the analysis of Chapter 3 as the correlation is calculated at each structural level of the model. Correlations between the scales are large, especially at the Type of Place and Household levels. From this it is possible to infer that there are broad similarities across the aggregated scales. There is, however, far less correlation at the individual level. This implies that even though higher-level results show consistency – such that households with high levels of wellbeing will have low levels of mental illness – the same is not true of the responses for individuals. Type of place has the largest correlation of 0.98, which effectively suggests that types of places with high levels of wellbeing will almost always have very low levels of mental illness. Region shows a markedly lower correlation of 0.45, the lowest across all levels. Despite these consistencies, at an individual level, which is the focus of the questionnaires, they are to some extent evaluating different underlying characteristics with the coefficient of 0.54 indicating a true predictive capacity of 29%.

Structural Level	Correlation Coefficient
Type of Place (5)	0.98
Region (4)	0.45
PSU (3)	0.65
Household (2)	0.83
Individual (1)	0.54

Table 4.2: Correlations between GHQ-12 residuals and SWEMWBS residuals across all structural levels of the base model.

Source: Calculations made by the author using data made available from Wave 1 of Understanding Society.

4.4.2 Relative importance of geographical scale for each response

The second research question looks at the relative importance of the higher-level scales in determining the two outcomes. Table 4.3 presents the VPCs for each structural level excluding the individual level, this section is exclusively concerned with higher-level relationships and thus involves models without any explanatory variables included.

As seen across both responses Household level captures the most variance. It is unsurprising perhaps that this is consistent across responses given that Household level has one of the highest correlations (Table 4.2). However, this is often true of lower levels in models, as they act as catch-all categories for remaining variance not attributed to higher levels (Tranmer and Steel, 2001). Despite this dominance of the household level it is very noticeable that non-household levels are far more important for SWE (13.8%) than the GHQ (6.7%). This suggests greater geographical patterning outside the household and individual for SWE than GHQ-12. Whilst the ordering of importance across both responses is consistent, the relative impact of the higher levels is notably greater for SWE, specifically Region and Type of Place.

When considering the question of the relative importance of functional or locational geography it seems evident that across both measures, the type of place matters more than the specific location. It should be noted that the classification of type of place are constructed based on demographic variables of the individuals in that area, thus they would be expected to borrow from the importance of the individual level quite heavily. A large proportion, nearly 9%, of non-individual variation in SWE is attributed to the *type* of PSU people live in, double that for the GHQ-12. The broader geographical regions in which people live have far less impact for GHQ-12 than for SWE, both absolutely and relative to the type of place level. This indicates that regional clustering in mental health is of little importance for the mentally ill, but of much greater importance for the gradient of mental health outside this bottom category. This reinforces the notion from Table 4.2 that the type of place has far greater predictive capacity in inferring mental state than the region.

It seems clear from this initial analysis that mental illness is notably less sensitive to geography than wellbeing. Without wishing to over interpret this, it seems that geography is more of a determining factor for those who are not experiencing the worst mental health.

Pre Demographic	***Higher level VPC (Percentage)***	
Structural Level	**GHQ**	**SWE**
Type of Place (4a)	4.4	8.8
Region (4b)	0.6	2.1
PSU (3)	1.6	2.9
Household (2)	93.3	86.2

Table 4.3: Higher Level Variance Partitioning Coefficients for GHQ-12 and SWEMWBS across all structural levels.

Individual level has been omitted from the calculations[20]. Source: Calculations by the author using data made available from Wave 1 of Understanding Society.

[20] When included in VPC calculations, the Individual Level accounted for 76.5% and 75.6% of the unexplained variance in GHQ-12 and SWE scores respectively.

4.4.3 Demographic Patterning of each response

Developing from this original variance-components model, Model 2 now goes on to include demographic predictors, with the effects of these being presented for each response below. All include demographic variables significantly reduced DIC[21], and were therefore all retained in the model. Graphics using predicted average values are used below to compare the fixed effects for predictor variables. All graphics are generated from a model with all other predictors included, but held at their mean, implying that other demographic predictors are all taken accounted for in each figure. The random part specification of the model affects what is being displayed as the average score is predicted for each nested level. Therefore, the predictions are for the typical person in a typical household in a typical PSU, simultaneously within a typical region and type of PSU. Thus, the predicted confidence intervals are not as tightly bound to the estimate as may be expected from a sample of this magnitude. It is also important to remember that it is the trends and not the absolute scores that are of interest in as absolute scores are not equivalent across measures.

4.4.3.1 Sex patterning

Figure 4.3 illustrates the differences between sexes for both mental health scales. Females score significantly higher for the GHQ-12, indicating worse mental health, which is consistent with the majority of prior research (Propper *et al.*, 2005; McLean *et al.*, 2011). However, no significant differences are seen in the SWE scores, with males and females on average scoring almost identically. Both scales have been investigated and confirmed to be free of gender bias (Goldberg *et al.*, 1997; Stewart-Brown *et al.*, 2009) so it is unlikely that this is the sole cause

[21] $P<0.05$ given added degrees of freedom.

of the sex differences between the two. It has been suggested that male stoicism, manifested as a reluctance to ask for or receive support, is a significant factor in observed gender differences in mental health (Murray *et al.*, 2008). This could be manifested by males being more likely to indicate distress via disagreeing with a positive statement than agreeing with a negative one. It is possible, therefore, that the exclusively positively worded items from the SWE remove this element of stoicism from the results of the GHQ-12. If this is the case, it raises concerns about observed sex differences and their relation to different response tendencies, rather than truly different symptoms.

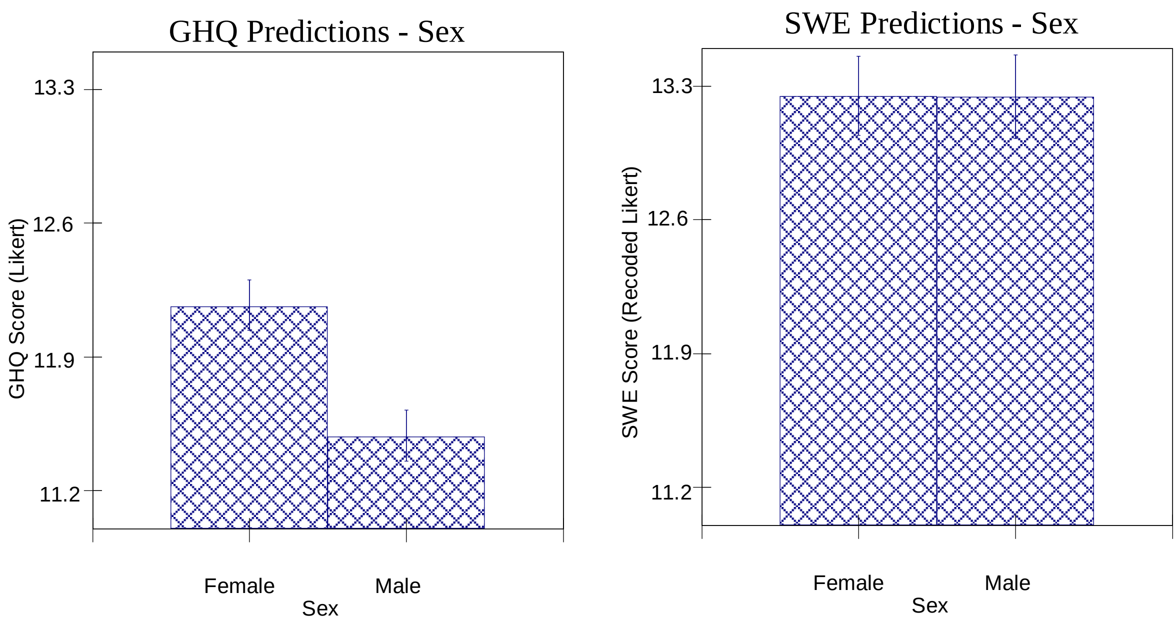


Figure 4.3: Predicted Scores for both responses stratified by sex, with coverage bounds indicated for each group.

4.4.3.2 Age Patterning

Age was included in the model as a quadratic function, allowing the investigation of the U-shaped curve evidenced in much of the literature. The effect of age can be seen in Figure 4.4, with similar age patterns being indicated across both scales, although the age of with the worst mental health predicted by the SWE is 2.2 years before the GHQ-12 predictions (45.8 to 48). This is consistent with previous research, supporting evidence of the mid-life peak suggested by Jorm et al. (2005) amongst others, and consistent with the age group with highest suicide rates in the UK (Samaritans, 2015). The SWE results indicate higher levels of distress in younger participants than GHQ-12, whereas in the GHQ the youngest and oldest age groups have very similar scores. It has been noted in previous research that stoicism is greater amongst older individuals, with younger individuals more likely to acknowledge depressive symptoms (Murray *et al.*, 2008), which would be consistent with SWE results, although not the GHQ-12.

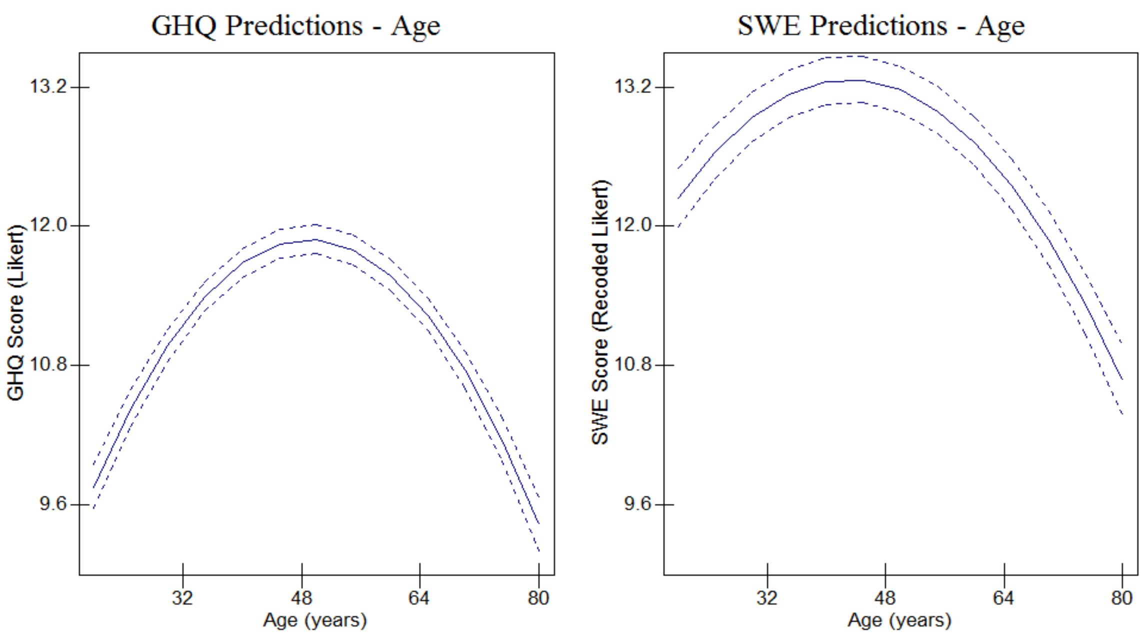


Figure 4.4: Predicted Scores for both responses against Age, with coverage bounds indicated for each group.

4.4.3.3 Ethnic Patterning

The effect of ethnicity[22] illustrated in Figure 4.5 shows similar trends across both scales, with overlapping CI's for all ethnic groups except for black people who score significantly lower across both scales. This contradicts a significant body of previous work (Lang *et al.*, 2011; Oguz, Merad and Snape, 2013). However, more recent work supports this notion, where it is suggested to be primarily due to black men having high levels of wellbeing (Stewart-Brown, Samaraweera, Taggart, N.-B. Kandala, *et al.*, 2015). Furthermore, there appear to be no significant differences between White, Mixed or Asian individuals for either of the responses, which again contradicts previous work which has found that Whites tend to enjoy the best mental health (e.g. Ross 2000). Propper et al. (2005) suggested an added complexity with respect to the homogeneity of an area, that being non-white has a protective effect in mixed ethnicity areas, but broadly conclude that contrary to results here, non-whites experience worse mental health.

[22] Understanding Society undertakes an ethnic minority boost sample to try and counteract the traditional under-reporting of ethnic minorities in panel studies. A more comprehensive overview of this is given in Chapter 5.

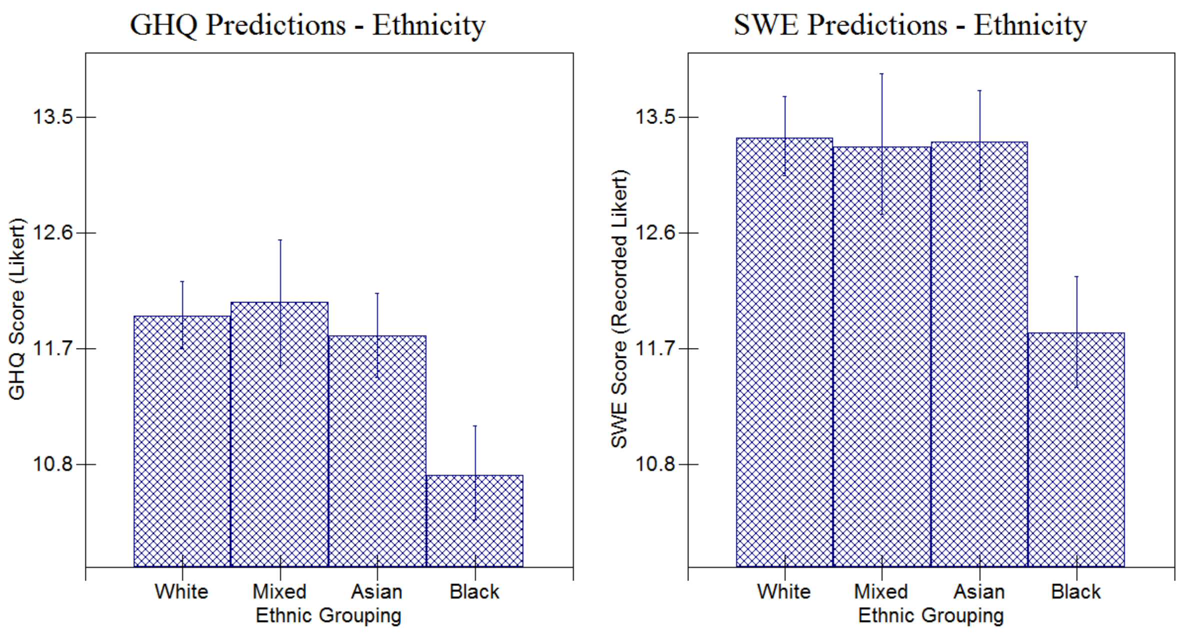


Figure 4.5: Predicted Scores for both responses stratified by Ethnic Grouping, with coverage bounds indicated for each group.

4.4.3.4 Patterning by Marital Status

The effect of marital status can be seen to be similar across measures in Figure 4.6, with the married enjoying the best mental health. This is consistent with previous research, with significantly lower odds of mental illness consistently found amongst the married population (Wade and Pevalin, 2004; Lindstrom and Rosvall, 2012). This effect has been suggested to insulate against poor mental health, but not extend to greater levels of high wellbeing (Stewart-Brown, Samaraweera, Taggart, N.-B. Kandala, *et al.*, 2015). Here, however the relationship seems similar for both. There looks to be a slight difference in singles versus SWD in GHQ-12 compared with SWE, but this falls short of significance.

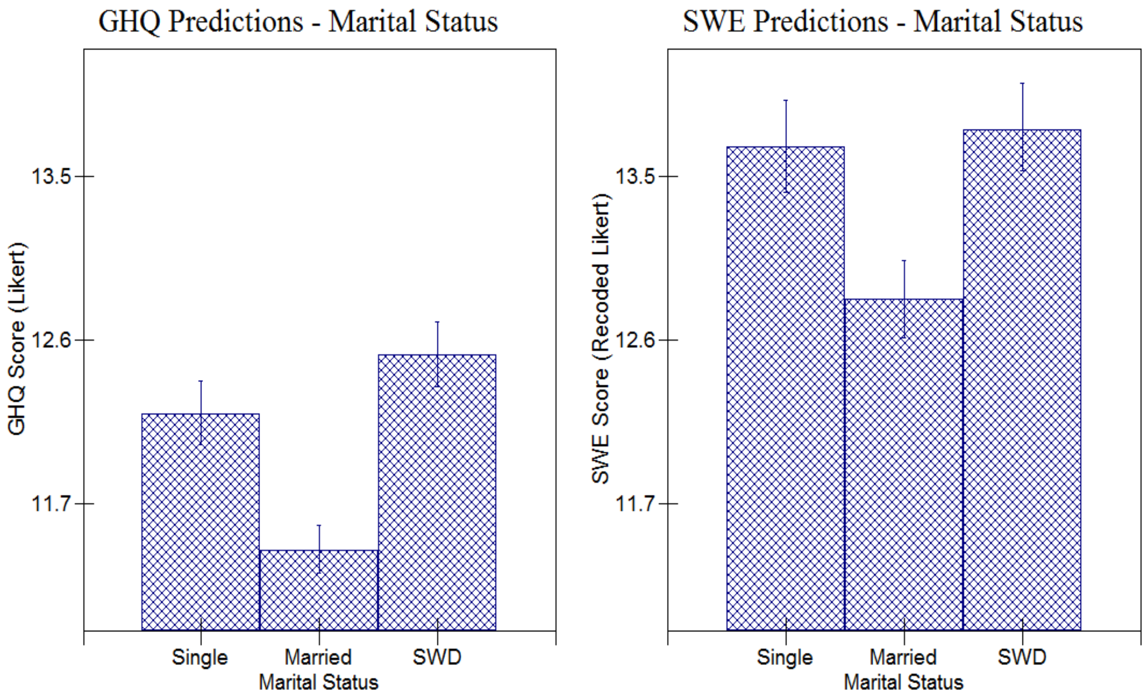


Figure 4.6: Predicted Scores for both responses stratified by Marital Status, with coverage bounds indicated for each group. SWD – Separated/Widowed/Divorced.

4.4.3.5 Patterning by Educational Attainment

Once socio-economic proxies have been included in the model, such as housing tenure and job classification – the differential effect of education, as Figure 4.7 shows, falls to insignificance for the GHQ-12. Previous research almost uniformly finds an association between education and mental health, with more educated individuals enjoying better mental health (Fryers *et al.*, 2005; Hu *et al.*, 2007). It is acknowledged that this relationship is likely to be complex due to education having high association with other known factors (Wilson and Oswald, 2005), several of which are accounted for in this model. There is not evidence for the diminishing returns on education suggested in a large volume of previous literature outside of the relationship seen for A-Levels in GHQ-12 responses. Interestingly there is still a clear dose-response relationship in SWE scores, with the more educated enjoying lower rates of poor mental health similar to that found by Stewart-Brown, *et al.*, (2015) who also find that this holds truer for women than men. This could suggest that the relationship with educational attainment is both stronger and more linear for mental wellbeing than for mental illness, with the mentally ill again being more demographically invariant.

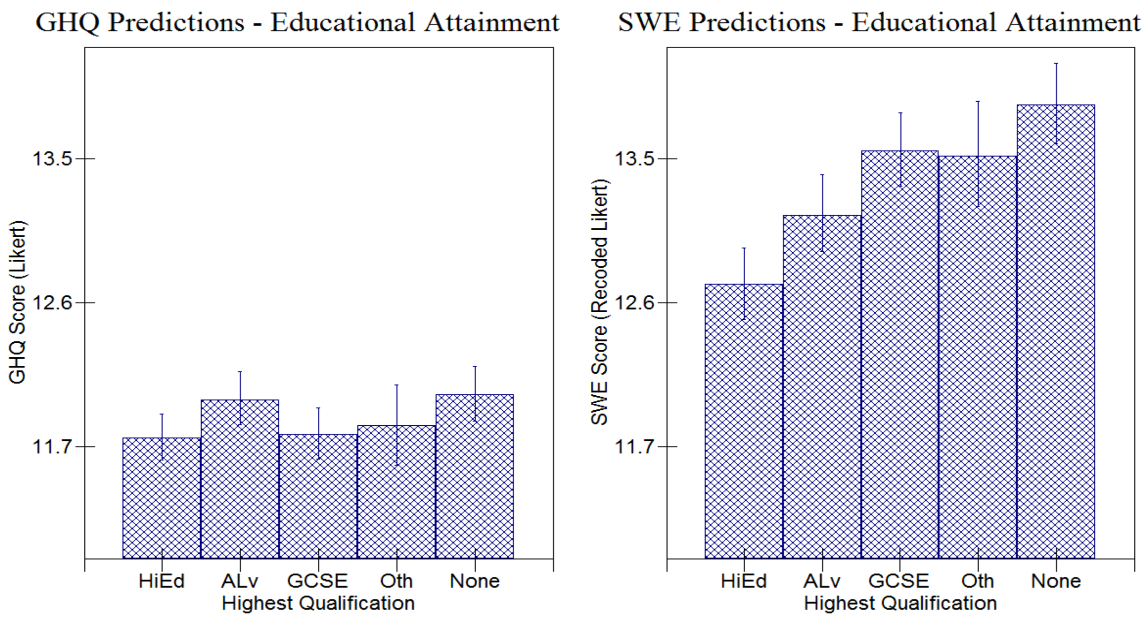


Figure 4.7 : Predicted Scores for both responses stratified by Highest Educational Qualification, with coverage bounds indicated for each group.
HiEd – Further Education (Degree Level), ALv – A Level, GCSE, Oth – Other Qualification, None

4.4.3.6 Patterning by Housing Tenure

Results for Housing Tenure as shown in Figure 4.8 are almost identical across the two scales – with those in social housing (local or housing authority rented) indicating the highest scores, and those who own their houses outright (OO) enjoying the best mental health. This corroborates previous work which finds consistently that renters experience worse mental health than homeowners (e.g. Meltzer et al. 2012). Interestingly there seems to be little difference in scores between those renting privately (PR) and those who own their house with a mortgage (OWM). The "other" category, consisting of non-rented or owned accommodation, also shows interestingly low scores but due to the low sample size (348), cannot be over interpreted.

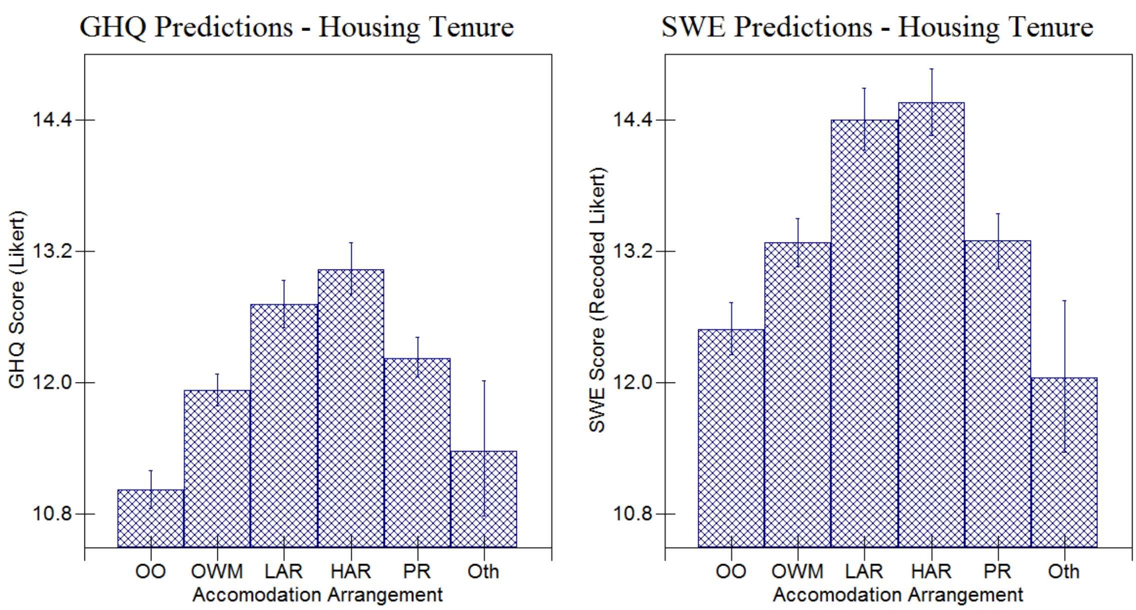


Figure 4.8: Predicted Scores for both responses stratified by Housing Tenure, with coverage bounds indicated for each group.
OO – Owned Outright, OWM – Owned with Mortgage, LAR – Local Authority Rented, HAR – Housing Authority Rented, PR – Privately Rented, Oth – Other.

4.4.3.7 Patterning by Employment Category

Job Classification seems to have quite a small impact outside of unemployment for GHQ-12, although Skilled Non-Manual, Partly-Skilled and Unskilled workers all experience worse mental health than those in employment. There is overwhelming consensus in the literature that unemployment is associated with poor mental health (Paul and Moser, 2009; van der Noordt *et al.*, 2014). It is acknowledged, however, that there are specific cases in which unemployment actually improves mental health (Ezzy, 1993), however this is not evidenced here, and is unlikely to be true when aggregated over a large scale as these instances are rare. The results for SWE show very similar results to the GHQ-12 with the largest effect being between the unemployed and the rest. It does seem that the differences between the employed groups are slightly greater for the SWE. For instance, notably higher responses for Skilled Non-Manual workers in SWE than GHQ look to suggest this type of work might have an effect on wellbeing which does not extend to mental illness.

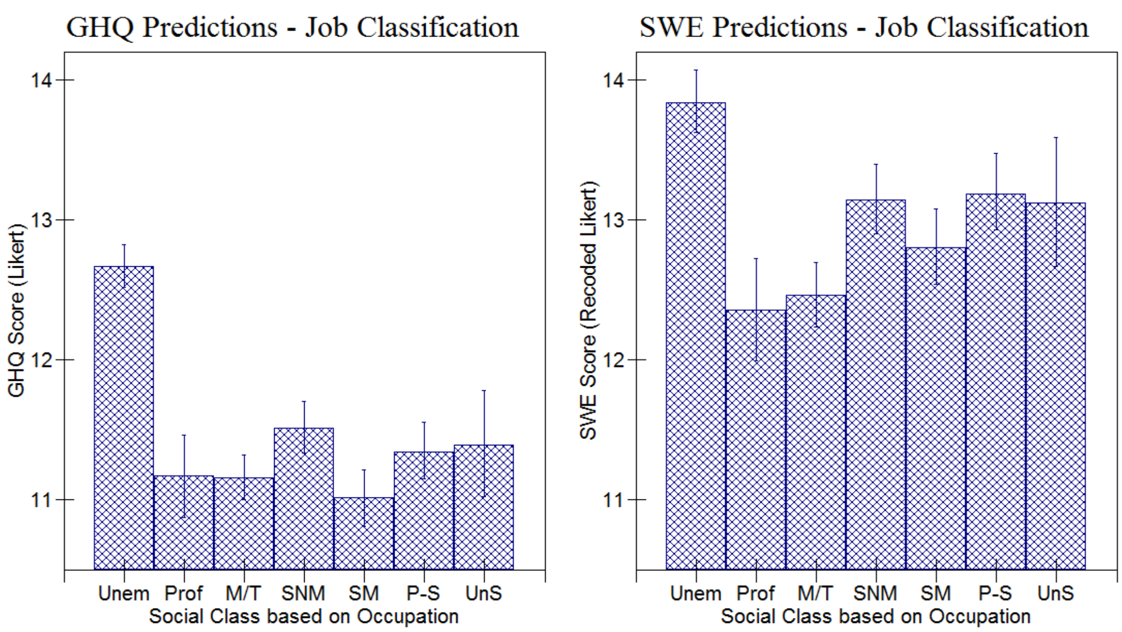


Figure 4.9: Predicted Scores for both responses stratified by Job Classification, with coverage bounds indicated for each group.
Unem – Unemployed, Prof – Professional, M/T – Managerial or Technical, SNM – Skilled Non-Manual, SM – Skilled Manual, P-S – Partly-Skilled, UnS – Unskilled.

4.4.4 Demographic attenuation of Area Relationships

We now return to the initial correlation and variance partitioning displayed in Tables 4.2 and 4.3 after controlling for these demographic variables. Table 4.4 gives the updated VPC estimates after controlling for demographic characteristics. The introduction of these demographic variables has greatest effect on the Type of Place VPC, reducing it substantially. It should be noted that the classifications are constructed based on demographic characteristics (Vickers and Rees, 2006), which has now been controlled for at the individual level. Therefore this reduction in VPC is expected, and should not be over interpreted. What this does suggest is that the majority of the differences between the *types of places* are compositional, rather than actual systemic differences between the types of place. More notable is the differential change in this level pre and post demographic controls. Between the scales the twofold difference remains, however within the scales, the drop in importance is much greater for the SWE. Again, this is likely to be mainly as a result of demographic construction of the Type of Place, however, given that this had a far greater effect previously, it is evident that SWE has a far greater capacity for demographic prediction, where GHQ-12 is more contextually independent. Again, if the assumption of GHQ-12 capturing illness and SWE capturing wellbeing holds, this supports the earlier assertion that the greatest effect of context is on those who are not experiencing the worst mental health.

Of further interest is the change in the strictly geographical levels. When factoring for demographic variables, there is nowhere near the reduction seen in the Type of Place level. In fact, for the GHQ-12, the relative importance of the locational levels increases marginally across the board. Again, this may be due to some of the uncaptured variance being previously attributed to the Type of Place level, but still illustrates the importance of geography. Similarly, SWE shows an increase in the relative importance of the Region-Level, suggesting that true differences between places were potentially being picked up by the type-of-place difference. Uncaptured variance in outcome at individual level will be expressed primarily at the most fine-grained spatial scale, so as expected, the household level gains importance, acting as a catch-all category for the newly explained variance.

Post Demographic Variable Inclusion	Higher Level VPC (Percentage)	
Structural Level	**GHQ**	**SWE**
Type of Place (4a)	0.3 (4.4)	0.6 (8.8)
Region (4b)	0.7 (0.6)	2.6 (2.1)
PSU (3)	1.7 (1.6)	2.3 (2.9)
Household (2)	97.3 (93.3)	94.5 (86.2)

Table 4.4: Higher Level Percentage VPCs after controlling for demographic characteristics in the model. Pre-demographic values given in parentheses.
Source: Calculations carried out by the author using data from Understanding Society Wave 1.

After introducing demographic variables there is a decline in correlation across all levels as seen in Table 4.5. The correlation at Type of Place level predictably falls, similarly to Table 4.3. Again this is assumed to be due to this level capturing some of the similarities associated with demographically similar people. However, it does not fall a great deal. Despite being very high initially, it still remains the second highest correlation level – suggesting, perhaps intuitively, that there is consensual experience of both positive wellbeing and reduced mental distress within similar types of place. There is little change in the correlation at region level after demographic controls, which is also intuitively reasonable as initial demographic similarities would be expected to be captured at levels 4a and 1. PSU level correlation falls, suggesting that neighbourhood similarities in mental health are somewhat a product of demographic similarities. Household and Individuals levels see very little and no difference respectively, despite these being the levels of the demographic controls. Whilst the residual variability at these levels fell as measured by the variance values in the model, the basic patterning of this proportional variance across the two levels remains the same, illustrated by the correlation. Thus it is clear that the relationship does not stay identical, the household and the type of place remain the levels which experience the most consensus between wellbeing and mental illness.

Structural Level	***Correlation Coefficient***
Type of Place (4a)	0.75 (0.98)
Region (4b)	0.42 (0.45)
PSU (3)	0.47 (0.65)
Household (2)	0.82 (0.83)
Individual (1)	0.54 (0.54)

Table 4.5: Correlation coefficients between the two responses across all spatial scales after factoring for demographic controls. Pre-demographic correlations are given in parentheses.
Source: Calculations carried out by the authors using data from Understanding Society Wave 1.

4.4.5 Modelling At-Risk Areas

So far, this chapter has been concerned investigating geography by summarising differences between higher level entities as variance terms. Here this variance is further analysed, examining specific differentials where it seems sensible to do so and identify the best and worst areas and types of areas to live in for mental health. These are presented via labelled pairwise residual plots, constructed after the inclusion of demographic controls.

Figure 4.10 illustrates the pairwise residuals from each response for each of the 53 types of place. The positive relationship between the two residuals reflects the 0.75 correlation observed in Table 4.5. Area type 4C1 “Prospering Suburbs – Semi-Detached” performs the worst on both scales, with area 4B1 “Prospering Suburbs – Older Families” performing best on both. This is surprising as the original construction of these categories was done in a stepwise manner, such that 4B and 4C subgroups are both within the larger “Supergroup 4” classification of “Prospering Suburbs”. This Prospering Suburbs category has the second highest mean property wealth of the 7 higher level groups (Daniel and Bright, 2011). This suggests that despite this assumed social homogeneity from similar classification based on the census, that there is notable dissimilarity in their experience of mental health. The worse mental health experienced by the “prospering semi-detached” areas could be to do with their relative property wealth, as it has been indicated that the 4C subgroup has notably lower than average property wealth than the Supergroup average (£160,445 compared with £223,443) (Daniel and Bright, 2011). Age and home ownership could also be partly responsible for the effect although these are controlled for in the construction of the areas (Vickers and Rees, 2006). This would imply that the effect suggested in Figures 4 and 8 could be the reason for the relative ill mental health of those in area 4C1. It seems that the better mental health experienced by older individuals, as illustrated in Figure 4, does not merely end at the age of the individual. Being in an area which is composed of other older and "prospering" people is also seemingly beneficial to the mental health of individuals living within it. Overall wealth across the groups also seems formative in an explanation, with 4b having the third highest overall average wealth (£713,978), and 4c the lowest within the Supergroup (£414,573) (Daniel and Bright, 2011). This heterogeneity within groupings seems to support some criticisms of the inherent variability captured within area classifications (Harris, Johnston and Burgess, 2007). Despite this clear patterning, the total degree of variation is not that large, with a range of 0.5 between the extremes in the Recoded SWE residuals, and 0.3 for the GHQ-12.

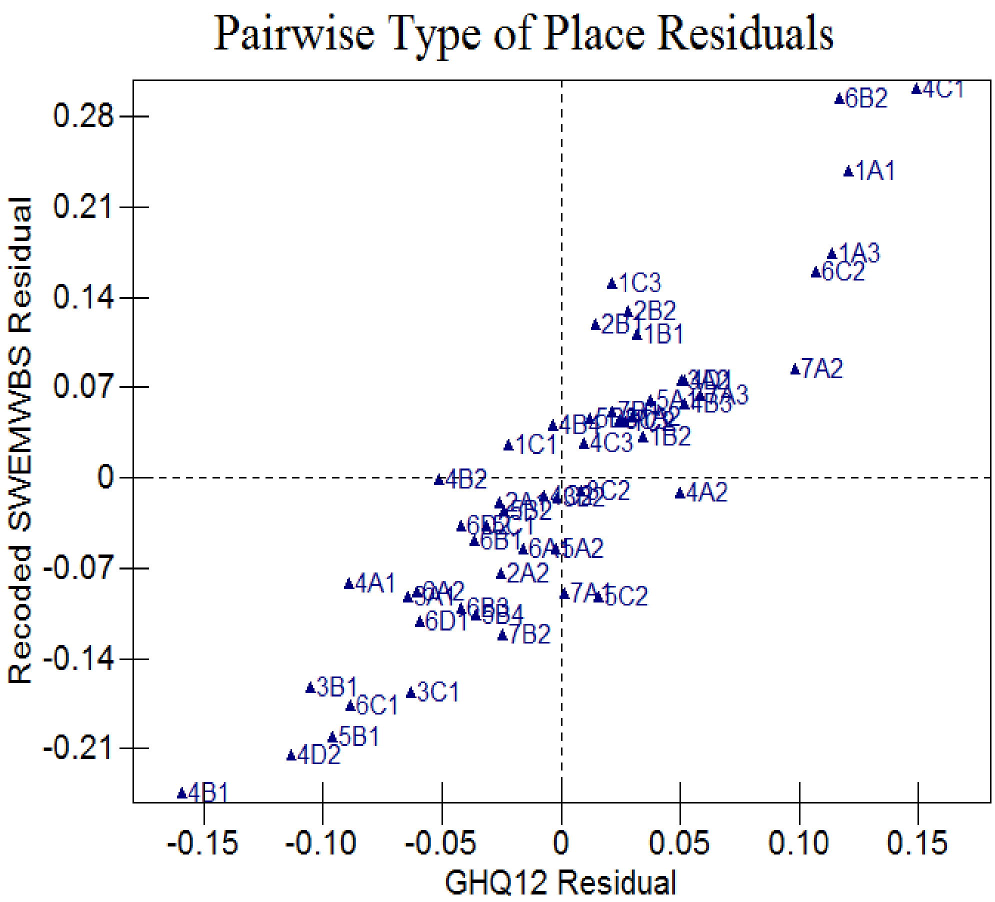


Figure 4.10: Modelled Pairwise Residual Plot of Type of Place Classifications (for full classification list see Vickers, 2006)

Figure 4.11 gives us the same type of graphical display but now the residuals are for UK regions. After factoring for demographic controls the region with worst mental health is the "West

Midlands Conurbation", scoring worst on both measures. This incorporates the city of Birmingham and surrounding conurbation, including Wolverhampton and Coventry (Jones, Johnston and Pattie, 1992). Northern Ireland experiences the best SWE results, and "Rest of Scotland" (the area excluding Strathclyde and the East Central region) enjoys the best GHQ-12 scores. Notably the best mental health across both measures is experienced in Scotland – with all regions excluding Strathclyde scoring lowest across measures. Previous research tends to paint Scotland as having poor mental health as a product of its socio-economic status. However, here this relationship is controlled for in the model itself, leaving just the raw effects of the area which seem beneficial. For a given socio-demographic status (as measured by the demographic variables in Table 4.1), individuals in Scotland experience the best mental health in the UK. Despite the positive correlation between residuals seen here, there are some which do not conform to the trend, for example Mersey scores slightly lower than average on the SWE but considerably higher than average on the GHQ-12. Furthermore, Northern Ireland scores the lowest on the wellness scale, however is almost exactly average on the GHQ-12 scale. This strongly suggests that researchers and policymakers should refrain from attempting to generalise from wellbeing to mental illness as there are clear and demonstrable differences even after factoring for demographic controls.

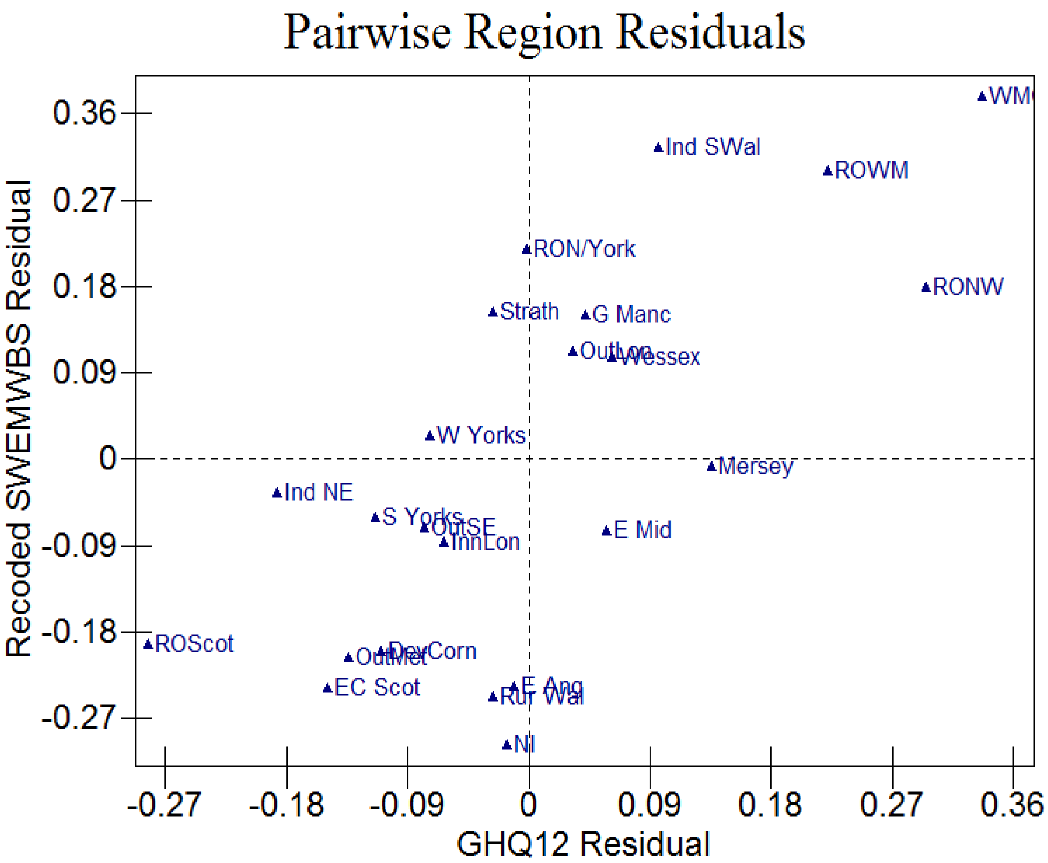


Figure 4.11: Modelled Pairwise Residual Plot for Higher Level Regions (For full classification information see Jones et al, 1992).[23]

[23] ROScot – *Rest of Scotland*, EC Scot – *Edinburgh and Central Scotland*, Ind NE – *Industrial Northeast*, S Yorks – *South Yorkshire*, OutSE – *Outer Southeast-* , InnLon – *Inner London*, OutMet – *Outer Metropolitan*, DevCorn – *Devon and Cornwall*, NI – *Northern Ireland*, E Ang – *East Anglia*, Rur Wal – *Rural wales*, W Yorks – *West Yorkshire*, Strath – *Strathclyde*, RON/York – *Rest of North and Yorkshire*, GManc – *Greater Manchester*, OutLon – *Outer London*, Wessex – *Wessex*, E Mid – *East Midlands*, Ind SWal – *Industrial South Wales*, Mersey – *Merseyside*, ROWM – *Rest of West Midlands*, RONW – *Rest of North-West*, WMid – *West Midlands Conurbation.*

4.4.6 For what age group does Geography matter most?

The final research question addresses whether geography is an equally important determinant of mental health for all, or whether there are specific groups who are especially sensitive to geographical effects. To investigate this final research question, a complex parameter is specified with respect to age, generating a level-specific variance estimate for age. This variance is then partitioned into that at the region level, and that at the type-of-area level, allowing us insight into; for which age geography is most important; and whether locational (region) or functional geography (type of place) is of greater importance. [24]

As seen in Figure 4.12 – the variance associated with the age term is greatest in youth and in old age across both measures and levels. This suggests that the times of greatest mental distress (middle age) are also the most geographically invariant. For the GHQ-12 both locational and functional geography seem to be similarly important, although is it noticeable that in middle age the functional geography retains more importance than the locational geography. The effect of geography on SWE results is much larger – having a far greater total variance associated over the lifecourse. This is particularly noticeable at the Region level variance function, which shows that for extreme ages scores can vary by almost 4 units, the equivalent of almost the entire range of the predicted scores for SWE for Age as a fixed effect. However, it is still the case that mental distress in mid-life is most consistent. In addition, for SWE there is much clearer indication of the importance of locational geography over and above that of the functional geography, being associated with nearly two times the variance of the Type of Place.

[24] This model is run without demographic controls other than age in order to isolate the raw relationship across levels.

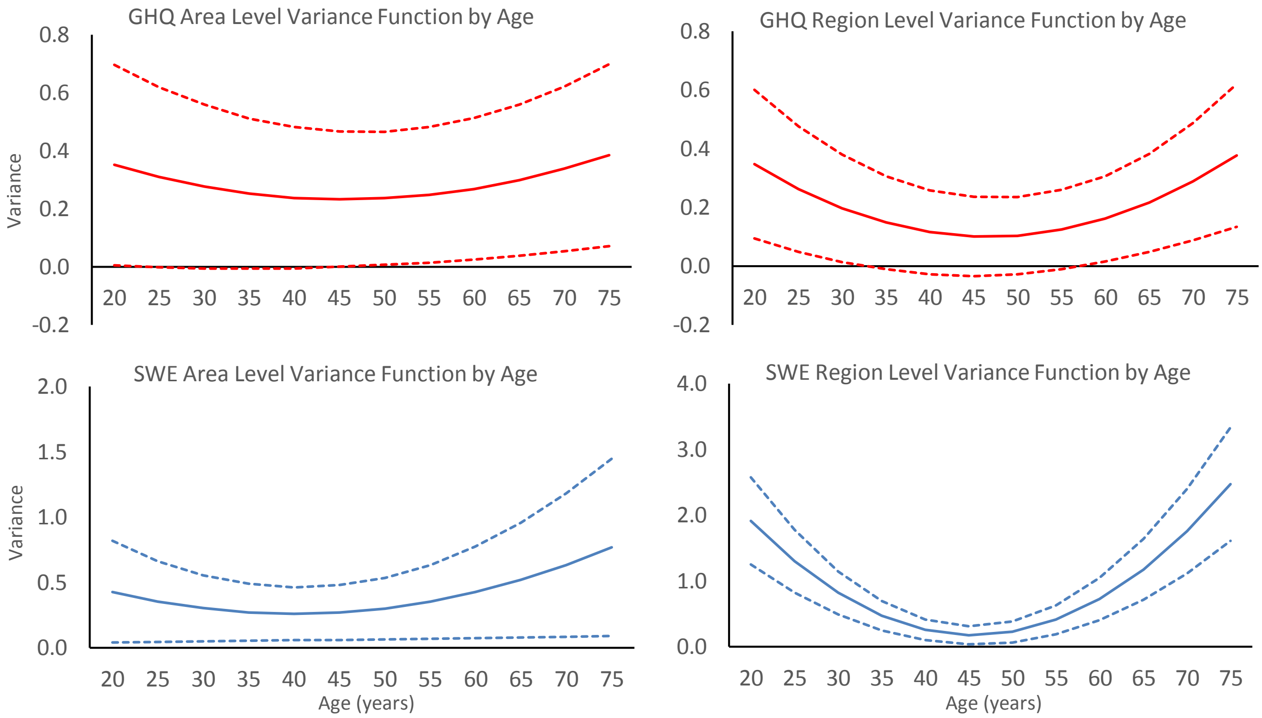


Figure 4.12: Variance Function Plots illustrating the Variance associated with different ages for both Region and Area Level, for both Responses. Top Row is GHQ-12, Bottom Row is SWE, Left Hand Side is Area-Type Level, Right Hand Side is Region Level.

Across both responses, geography matters the most for the young and the elderly, with greatest variation present for these groups. The peak of mental distress, seen in Figure 4 above, in middle age is also the age for which there is least variation, suggesting that this is consistent across contexts. This also supports the suggestion from Table 4.2 that groups experiencing the worst mental health seem to do far more uniformly with respect to geography than those outside this band. This is most notable in the Region Level variance function for SWE. It is also noteworthy that there is little difference between the variability of the young and middle aged with respect to Type of Place, and that this all seems to be expressed at the region-level.

4.5 Conclusions

This chapter investigated two mental health responses and their distribution across the UK, both with respect to demographic and geographical variables. Furthermore, it investigated the relative importance of the type of area compared with strict location as part of a 5-level spatial structure, specifying a cross-classified model which simultaneously modelled both the specific hierarchical location and the area classification. This allowed the extension of variance partitioning beyond the strict hierarchy which typically confines multilevel models.

A large degree of consistency was found between predictions for the GHQ-12 and the SWE, but that this is primarily at higher spatial scales, with individual level responses having relatively low correlation across the two measures. Across spatial scales the most important scale outside the individual is found to be the household. Outside of the household, the most important level is the type of place, suggesting that functional geography has greater impact than broad locational geography in determining mental health outcomes. Furthermore, initial higher level VPCs indicate that GHQ-12 responses are more geographically invariant than SWE outcomes, which suggests a greater geography to wellness than illness. Despite these aggregated results showing high degrees of consistency, caution should be exercised when generalising as this is not true of the individual level. The individual level correlation of 0.54 shows that individuals with a high level of wellbeing will not necessarily be free of mental illness, and vice versa.

There were large differences between different demographic groups, with age and employment both having notably large effects, with a midlife peak in mental distress discernible across both responses. Interestingly for employment, the GHQ-12 exhibits a broad binary in health discrepancies for employment status, between the employed and not. The SWE however, illustrates a more gradual change across different types of work, suggesting that concentrating on mental illness alone misses some key effect of employment type. Additionally, the effects of sex and education were different across responses, no difference across sexes for SWE outcomes, and education having no effect for the GHQ-12.

The initial importance of functional geography was attenuated to an extent by the inclusion of demographic variables, which are suggested to be a manifestation of inherent compositional effects in area classifications. Despite this, some importance of the area classification remained even after factoring for individual characteristics, suggesting that there is an effect associated with the type of area people live in, over and above the specific location in which they reside. As suggested above it seems that neighbourhood or regional effects are far greater for wellness than for illness measurement, with the very ill seeming far more independent of context. Perhaps in future study, this should be addressed as two separate questions, with different geographical patterning predicted for illness and wellness.

Moreover, Variance Partitioning Coefficients suggested a greater geographical sensitivity of wellness compared with that of mental illness. As demographic variables were included, the importance of all levels decreased for the GHQ-12, where they retained far greater importance for the SWE. Whilst this may tie into some causal arguments in the literature we do not assert any claims as to the causality of this relationship here. The household is suggested to be an important level for social policy intervention as households with a high degree of wellbeing are very likely to have a low degree of mental illness, with the highest correlation across spatial levels. Targeted funding here may do more to improve mental health across the whole wellness spectrum, rather than focusing specifically on the ill. It is also notable that individual level correlation remained the same after factoring for demographic characteristics, remaining at 0.54, implying that at an individual level it is still misguided to think of low wellbeing and mental illness as interchangeable.

Once the initial age relationship was unpacked further it became clear that not only are the middle aged the most mentally distressed, but they are also the least variable in their mental health with respect to geography. The mental health of the young and the elderly is far more sensitive to geography than that of the middle aged, who have evidenced the worst mental health in Figure 4. This supported the suggestion that geography matters more for those who are not at the low end of the mental health spectrum, and that those who are suffering more severely are likely to do so regardless of geographical context.

The UK is in need of a more developed mental health infrastructure, as evidenced by suicide now being the largest killer amongst men below 45 (Samaritans, 2015). To effectively target those who are most at risk, there needs to be considerable work done on the most efficient use of public funds. Here we present evidence that perhaps more simplistic understandings of the broad associations between financial distress and mental health that have often been relied upon, do not give us the whole picture. Whilst this is likely to be the best strategy for investment in mental health nationwide, it should be considered that there are likely to be individuals at risk of mental distress regardless of context across the nation. Perhaps evidence of this more homogenous mental health experience for the very ill, across demographic divides is likely to aid in the removal of mental illness stigma nationwide.

5 LONGITUDINAL ANALYSIS OF THE TEMPORAL CONSISTENCY, DEMOGRAPHIC PATTERNING AND DIFFERENTIAL TRENDS OF FOUR MENTAL HEALTH CONSTRUCTS AS CAPTURED BY THE GHQ-12.

5.1 Chronicity in Mental Health

Based on figures published by the Office for National Statistics (ONS) and NHS reports based on the Adult Psychiatric Morbidity Survey (APMS), many charities and mental health advocates highlight that 1 in 4 adults will experience a mental health issue in any given year (Mind, 2016, Mental Health Foundation, 2017). Whilst these statistics are useful for raising awareness of the scale of problems that mental health poses to the UK population, there is little in these reports discussing the degree to which these conditions are stable over time within individuals. As discussed in previous chapters, mental health measurement was initially far more psychosis-oriented than contemporary "mental health" measurement. In the psychological literature, psychoses such as schizophrenia are often characterised as *chronic*, meaning they are accompanied by a diagnosis that lasts for a significant number of years, if not life (Breier *et al.*, 1991; Fioravanti *et al.*, 2005; Park *et al.*, 2017). Conversely, more common mood disorders such as anxiety and depression are traditionally thought of as episodic, or remitting (Keller *et al.*, 2000), and are most commonly discussed in terms of "periods of depression" over the life course in combination with discussion of recovery (e.g. Leamy *et al.*, 2011; McManus *et al.*, 2016). Whilst the ubiquity of chronicity in schizophrenia has been challenged (Harding *et al.*, 1987), psychosis-oriented illnesses are characterised by chronicity to a much greater degree than mood disorders.

5.1.1 Mental Health Stability

Social science research has tended to focus more on mood disorders than psychoses due to their greater disease burden (World Health Organization, 2013). As such, temporal consistency in mental health as considered in quantitative social science has yet to receive this degree of attention. A large body of research has investigated the longitudinal trends in mental health and suggested a global trend towards increasing numbers of individuals with depressive disorders (Levinson *et al.*, 2010; Lépine and Briley, 2011; World Health Organization, 2013). This trend of worsening mental health has also been evidenced in a UK context across both clinical outcomes and self-completion mental health screenings (Barr *et al.*, 2012; Katikireddi, Niedzwiedz and Popham, 2012; Spence *et al.*, 2014; Barr, Kinderman and Whitehead, 2015; Collishaw, 2015). These investigations - regardless of geographical context - have overwhelmingly focused on longitudinal development at the aggregate scale, only identifying if there are greater raw numbers of "ill" individuals in a given year (e.g. Weich *et al.*, 2011; McManus *et al.*, 2016). Whilst this is undoubtedly important and relevant, the conclusions drawn from this research can go no further than saying that in a given year there were a greater number of ill individuals; they cannot characterise changes in mental health within individuals longitudinally. Little consideration has been given in quantitative panel studies as to whether individuals experiencing mental distress in a given year are the same individuals as the year before.

5.1.2 Why care about Temporal Consistency?

The temporal consistency of individual mental health is an issue of great concern in understanding how to approach the issue of mental health at the population level (Pevalin, 2000; Lorant *et al.*, 2014; Steptoe, Deaton and Stone, 2015). Whether individuals who become ill then stay ill has implications for their treatment and for the treatment of mental health at the population level. This knowledge could be used to inform whether variability is higher between-individuals across -timesteps, indicating it is individuals who are at risk, or whether variability is higher within-individuals between-timesteps, indicating periods of distress, rather than consistently distressed individuals. Imagine that it is categorically shown that there is greater variability in mental health between occasions (having taken account of the individual) than there is between individuals. This could imply that temporally-variant, environmental distressors which influence the mental health of an individual at a given time are important, regardless of the specific individual. This could lead to policy recommendations targeted at improving the factors that are consistent across all individuals but vary with respect to time, such as mental health awareness, education more broadly, unemployment, inequality or deprivation. If, conversely, it was the case that the individual level is where the greatest variation lies, then it would suggest that mental health risk is consistent across individuals through time. This could be used to justify policy focussed on improving the circumstances of individuals via greater individual support and care, and continuation of care similar to other limiting, long-term illness. More simply this can be thought of as whether, when dealing with mental health, treatment should be framed in terms of “unhealthy individuals” or “individuals going through an unhealthy period”.

5.1.3 Uniform Consistency?

The notion of temporal consistency in mental health disorders is commonly discussed as “chronicity” or “remittance” in psychological literature. As discussed above, this has not been formally investigated in quantitative social literature. Furthermore, as outlined in previous chapters, there are issues with the measurement metrics commonly used to ascribe mental functioning. Chapters 2 and 3 illustrated the degree of dissimilarity within the GHQ-12 and the SWE and demonstrated the analytical limitations of assuming unidimensionality. Given that the cross-sectional results of these questionnaires were demonstrated to be patterned by substantively and empirically different underpinning constructs, it seems short-sighted to assume that the longitudinal trends in a composite metric will be uniform across all of its constituent elements. Multidimensionality must be evaluated at a longitudinal as well as cross-sectional level. Moreover, it is of substantive interest to see whether the structure of this multidimensionality is consistent both within and between individuals. For instance, it is possible to imagine a scenario where an individual with consistently low self-worth may find it harder to function socially through time. This scenario would not, however, necessarily imply that if there is a particular *time-period* where individuals struggle with low self-worth at a population level, that this will be associated with low social functioning at the population-level. It is thus important to consider a theoretical model which does not assume uniformity in the processes affecting distressed individuals and distressed time-periods.

5.1.4 Chapter Aims

This chapter seeks to address these gaps in understanding and analyse to what degree mental state is consistent within an individual or varies within individuals over time. This is not investigated in a unidimensional form but instead for each of the underpinning constructs of the GHQ-12, modelling the responses simultaneously over time. In specifying a sufficiently complex model to address this, it will also facilitate analysis of whether the demographic patterning illustrated in Chapter 4 is consistent over both; time, considering the following 6 waves from the 2009 wave of Understanding Society; and the latent GHQ-12 constructs, not simply the summed scores. Finally, it goes on to discuss to what degree trajectories in mental health experience are consistent across the population over the 6-wave, 8-year period. It will also characterise these dissimilarities by demographics, looking at different trends between groups.

5.1.5 Research Aims.

Four broad research aims present themselves:

1. Over this 6-year period, how temporally stable are these constructs between areas, between individuals and within individuals? Are individuals consistent in displaying mental distress as defined by each of the four constructs?

2. How correlated are each of the four mental health constructs, at each structural level?

3. Averaging over the 6-year period, which demographic groups are most vulnerable to mental distress? Do these at-risk demographics differ across each of the four constructs?

4. Having looked at aggregate differences over the entire time period, are different demographic groups experiencing different trajectories in mental health in the period 2009-2014?

Ultimately, this final chapter draws on the increased methodological complexity gained from the adoption of a multidimensional approach to mental health measurement outlined in Chapter 2, and combines it with a multivariate, longitudinal, multilevel model, developing on the strategies outlined in Chapter 4. An exploratory model-building procedure is adopted, seeking to identify the best explanation of the mental health trends in the UK over the period 2009-2016. Data for the GHQ-12 is taken from 6 waves of Understanding Society. This data is then decomposed into the four constructs developed in Chapter 2, each of which is specified as a response. Complex variance is specified for each of the responses, with covariance specified between them at each structural level. This allows the estimation of the degree of consistency across PSUs, within PSUs between individuals and crucially within individuals between occasions. This addresses the degree to which mental health varies over time within an individual for each of the four responses. Separate effects are then estimated for a number of demographic predictors for each response. This specification allows identification of broad longitudinal trends and different at-risk groups for mental health having taken account of the year of measurement. Finally, interactions between demographics and wave-year are introduced into the model. This allows the estimation of differential trends for different groups over the time period, allowing the identification of groups experiencing disproportionately worsening mental health. It concludes with a comprehensive overview of mental health as captured by the GHQ-12 over the period from 2009-2016, identifying some of the more interesting findings from the research.

5.2 Data

The data here are taken from the first 6 Waves of Understanding Society. It comprises 239,372 observations, collected from a representative sample of 71,161 individuals over 6 measurement-occasions. The waves run sequentially from 2009 until 2014, however data collection for each wave constitutes 24 monthly samples such that each year label indicates data collected in the two years following it (i.e. the “2009” wave is released in late 2011 and so on (McFall, 2011)). As in Chapter 4, this chapter exploits the clustered nature of the data collection, using Primary Sampling Units (PSUs) as a geographical identifier to help identify area-level differences in mental health.

5.2.1 Attrition in Panel Studies

Previous chapters have not considered longitudinal changes, thus attrition in the panel population between waves was not an issue. However, given the consideration of temporal trends, how consistent the panel is at each wave is now a matter of interest and importance. Understanding Society attempts to maintain Original Sample Members (OSMs) from Wave 1 as long as they do not leave the UK. Individuals joining the household of an OSM become Temporary Study Members (TSMs) (Understanding Society, 2017). A male TSM who fathers a child with a female OSM becomes a Permanent Study Member (PSM), meaning they are then followed at every wave similarly to an OSM, even if they leave the household, unlike a TSM (*ibid*.).

Despite this, attrition may occur where households do not respond, refuse to be assessed again, or for another reason have their response not included. Furthermore, participants within an otherwise assenting household may still refuse to respond, or not be able to respond in the manner necessary for the study. This is not an issue in the large majority of cases, for example

77.1% of the interviewees eligible to be in Wave 2 who were present in Wave 1 gave full interview responses. The panel grows at each wave with the inclusion of new TSMs and sample boosts targeted at specific groups of interest. These include the Ethnic Minority Boost Sample (EMBS) and later the Immigrant and Ethnic Minority Boost Sample (IEMBS) at Wave 6. Additionally, from Wave 2, the individuals who were in the original British Household Panel Survey (BHPS), from which Understanding Society was developed, were included in the dataset. The introduction of these new individuals at each wave broadly account for the numerical element of attrition rates, keeping raw responses numbers relatively stable. Table 5.1 gives the year on year development of retention rates among the General Population Sample (GPS), EMBS and Former-BHPS respondents between waves.[25]

[25]Note, % retention is percentage giving full interview and is calculated as % of eligible previous panel members at next wave, not % of total previous wave respondents. Also note these values do not include additional individuals from original BHPS Sample who are included in the total data.

Population by Wave	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
GPS Total Interviewee Population	43,674	34,151	30,685	28,984	27,533	25,064
GPS Percentage consistent from Previous Wave	*N/A*	77.1%	78.6%	82.7%	85.2%	84.6%
EMBS Total Interviewee Population	7,320	6065	4,432	4,236	3,961	8,007[26]
EMBS percentage consistent from Previous Wave	*N/A*	67.3%	69.4%	74.6%	78.6%	72.9% *(of 3,549)*
Former-BHPS Total Interviewee Population	*N/A*	12,461	11,726	10,853	10,327	10,306
Former-BHPS Percentage consistent from previous wave	*N/A*	*N/A*	80.4%	80.2%	82.1%	93.1%
Total GHQ-12 Responses	39,700	43,446	40,612	38,852	37,196	39,566
Total GHQ-12 Responses to all Previous Waves	39,700	27,182	20,684	17,540	15,473	13,152

Table 5.1: Development of Panel Composition over the 6 Waves.

The initial wave in 2009 had 39700 individuals complete the full GHQ-12, as outlined in Chapters 2 and 3. However, not all these individuals remain in the cohort for the full 6 waves. This can present a problem for obtaining generalisable conclusions if these individuals drop out of the panel not-at-random. The seventh row of the table gives the numbers of individuals completing the full GHQ-12 at each wave. These are the individuals who make up the data analysed in this chapter. The following row gives the number of individuals responding to the full GHQ-12 in each wave who have responded in all previous waves[27]. If the individuals

[26] Wave 6 saw the introduction of the Immigrant and Ethnic Minority Boost sample, comprising another 4,458 individuals from migrant and minority ethnic backgrounds. IEMBS is not included in the calculation for % consistency.

[27] Note, this overestimates the effect of attrition, as there are a number of individuals who leave the panel for at least one wave and rejoin at a later date.

dropping out of the study at each wave are patterned demographically, then this has the potential to bias results towards those individuals who remain in the panel.

5.2.2 Disproportionate Demographic Attrition

The potential bias introduced by a non-representative sample has been argued to not matter under certain circumstances (Goldstein *et al.*, 2015). This includes when investigating research questions which pertain to a population which differs from the general population of a country at a given moment (Ebrahim and Smith, 2013; Rothman, Gallacher and Hatch, 2013).

However, even if this effect is small, it is still a useful and necessary exercise to formally investigate this attrition and the resulting representativeness of those responding at each wave. To do so, the demographic characteristics of the initial 39,700 GHQ-12 respondents is contrasted with the 13,152 individuals who responded to the GHQ-12 in every wave. The results of this are presented below in Table 5.2.

	Wave 1 GHQ-12 respondents		All-Wave GHQ-12 Respondents		Percentage Difference
	Frequency	Row %	Frequency	Row %	
Sex					
Female	22269	56.09	7627	57.99	1.9
Male	17431	43.91	5525	42.01	-1.9
Ethnicity					
Asian	3795	9.56	632	4.81	-4.75
Black	1837	4.63	302	2.30	-2.33
Mixed	875	2.20	219	1.66	-0.54
White	33166	83.6	11990	91.23	7.63
Highest Educational Qualification					
Higher Education	14,102	35.31	5,547	42.19	6.88
A-Level	7,707	19.42	2,454	18.66	-0.76
GCSE	7,960	20.06	2,573	19.57	-0.49
Other	4,063	10.24	1,273	9.68	-0.56
None	5,941	14.97	1,301	9.90	-5.07
Employment Classification					
Professional	1,389	3.51	552	4.21	0.7
Managerial/Technical	7,908	19.99	3,148	24.01	4.02
Skilled Non-Manual	4,866	12.30	1,729	13.19	0.89
Skilled Manual	4,145	10.48	1,394	10.63	0.15
Partly-Skilled	3,128	7.91	959	7.31	-0.6
Unskilled	819	2.07	219	1.67	-0.4
Unemployed	17,302	43.74	5,111	38.98	-4.76

Table 5.2: Demographic differences at baseline between those who participated in Wave 1 and those who participated in every wave.

Table 5.2 displays the differences in demographic characteristics between those that provided full GHQ-12 responses in wave 1 of Understanding Society, and those who completed the full GHQ-12 in all 6 waves. It is a slight bias towards females over the course of the study, as well as towards whites, the highly educated, and those in managerial positions. This is consistent with many longitudinal studies such as the Avon Longitudinal Study (ALSPAC) which find it hard to retain respondents with lower educational attainment (Boyd *et al.*, 2013). Relative to these studies, the demographic changes seen over the course of the panel here are small but longitudinal conclusions drawn in this chapter must be sensitive to the possibility that they are biased toward this slightly more unrepresentative cohort.

The principal concern of this chapter is in the temporal consistency of mental health. As such it is important that an understanding of the consistency of longitudinal respondents is established. Table 5.3 provides the numbers of waves participated in by each individual in the study to establish to what degree each study respondent contributes to the longitudinal and cross-sectional elements of the analysis.

Waves Responded To	**Frequency**	**Percentage**
1	19,288	27.10
2	9,908	13.92
3	7,332	10.30
4	8,045	11.31
5	13,436	18.88
6	13,152	18.48

Table 5.3: Breakdown of the number of sequential waves responded to by the 71,161 respondents throughout the observation period.

Table 5.3 shows that of the total 71,161 individuals included in the longitudinal sample 72.9% of them are contributing to the estimation of longitudinal trends over the observation period. That gives a true longitudinal sample of 51,783. The remaining 27.1% contribute to the precision in estimation of between-individual effects, including the robust estimation of demographic patterning, but do not contribute to the longitudinal component of the investigation. This gives this chapter remarkable statistical power in estimating longitudinal geographical and demographic effects.

Having established the representativeness of the longitudinal component of the dataset, it is important to now see the demographic and geographical structure of the respondents. The composition of this full, compiled data, along with raw proportions for each of the selected demographic predictors is given below in Table 5.4.

DATA STRUCTURE		
Structural Variables		
Primary Sampling Units	8993 Unique Units	
Individuals	68,545 Unique Individuals	
Observations	239,372 Observations	
Years	2009 16.6% [39700] 2011 17.0% [40612] 2013 15.5% [37196]	2010 18.1% [43446] 2012 16.2% [38852] 2014 16.5% [39566]
Response Variables		
Lowered Self Worth	Mean = 2.455, SD = 2.078	Min = 0, Max = 10
Social Dysfunction	Mean = 3.586, SD = 1.252	Min = 0, Max = 10
Stress	Mean = 3.121, SD = 1.933	Min = 0, Max = 10
Emotional Coping	Mean = 4.084, SD = 1.226	Min = 0, Max = 10
Predictor Variables		
Age (continuous)	In years (in 2009) Mean = 47.020, SD = 18.201, Min = 9, Max = 102	
Categorical age (AgeCat)[28]	Categorical Age, age in 2009 <25 8.7% [20831] 35-44 17.1% [41017] 55-64 16.2% [38785] 75-84 8.1% [19470]	 25-34 14.1% [33799] 45-54 18.8% [44891] 65-74 14.0% [33423] 85+ 3.0% [7139]
Sex	FEMALE 55.8% [133590]	Male 44.2% [105780]
Highest Educational Qualification	HIGHER EDUCATION (HiEd) 34.9% [83614] GCSE or Equivalent (GCSE) 20.9% [50147] No Qualification (NoQual) 12.8% [30557]	A Level Equivalent (ALv) 20.9% [50059] Other Qualification (Other) 9.4% [22611]
Marital Status	Single Never Married (Single) 22.5% [53769] Separated, Widowed or Divorced (SWD) 14.1% [33771]	MARRIED OR CIVIL PARTNERSHIP (Mar) 51.7% [123811] Living as a Couple (LaC) 11.5% [27644]
Ethnicity	WHITE 84.9% [203187] Asian 7.9% [18965]	Mixed 2.4% [5640] Black 3.8% [9116]
Job Classification by Registrar General	Professional (Prof) 3.6% [8566] Skilled Non-Manual (SNM) 12.4% [29700] Partly Skilled (P-S) 8.3% [19783] UNEMPLOYED (Unemp) 42.1% [100833]	Managerial/Technical (ManTec) 20.6% [49377] Skilled Manual (SM) 10.5% [25043] Unskilled (UnS) 2.0% [4757]

Table 5.4: Breakdown of the data modelled in this Chapter. Distributional Statistics are given for continuous variables. For categorical variables the base category is emboldened, capitalised and italicised. Percentages may not add to 100 due to rounding differences. Figures in square brackets give raw sample sizes for each category.

[28] Note: Categorical Age is only included for trend construction, in which continuous age is removed from the model, this is the only model for which Categorical Age is used.

5.2.3 Constructing Meaningful Response Variables:

Also included in Table 5.4 are the summary statistics for the four response variables. Meaningful and interpretable responses require two key characteristics. Firstly, the responses must reflect the underpinning dimensions of the GHQ-12 which were identified in Chapter 2. This structure was generated from the initial wave of the Understanding Society dataset, and so in generalising from this wave the responses necessarily assume the dimensions of mental health are consistent over the six waves. Secondly, the responses must be comparable both between constructs, and between waves. This requires a decision on standardisation. Given the interest in modelling potential changes in demographic groups, the object of interest is change relative to the population over the six waves. Thus, the standardisation needs to be carried out over the entire dataset.

To achieve the first of these characteristics the responses are initially constructed from raw GHQ-12 scores for each item. These scores are then simply multiplied by the associated loading as seen in Table 2.4 in Chapter 2[29]. They are subsequently summed based on the construct they comprise, giving four raw response variables. Due to differences in loadings, and differences in the numbers of constituent items for each construct, these raw constructs are not comparable with each other, and have very different scales.

---

[29] The specific formulae are as follows:
Lowered Self Worth =
(0.548*'GHQB')+(0.548*'GHQE')+(0.656*'GHQF')+(0.684*'GHQI')+(0.924*'GHQJ')+(0.854*'GHQK')
Social Dysfunction =
(0.627*'GHQA')+(0.619*'GHQC')+(0.856*'GHQD')+(0.554*'GHQG')+(0.732*'GHQH')+(0.514*'GHQL')
Stress = (0.229*'GHQA')+(0.464*'GHQB')+(0.596*'GHQE')+(0.339*'GHQF')+(0.198*'GHQG')+(0.269*'GHQI')
Emotional Coping = (-0.233*'GHQD')+(0.176*'GHQG')+(0.190*'GHQI')+(0.469*'GHQL')

This gives the motivation for the second key characteristic. For effect sizes to be comparable across responses, the responses need to be standardised. This standardisation not only needs to generate comparable units across responses, but also needs to be comparable longitudinally over the six waves. To do this, typically each response would be standardised into a normalised score (often termed a normit score), which transforms the ranking of each observation into an explicitly normally distributed response with a mean of zero and a variance of one. This approach is undertaken here, however there is one key difference. The rankings are not continuous, there are a finite number of possible scores that can be achieved by respondents, given the four possible responses to each item. This leads to some responses having somewhat non-normal distributions, with some responses showing skew around this zero-mean.

The key assumption of normality in the model is for the residuals, not the observed responses and as predictors are included it can be expected that the conditional distribution will become more normal. Moreover, Bell et al (in press) demonstrate with a simulation study (in line with previous research articles) that mis-specifying truly non-Normal effects as Normally distributed can slightly bias variance/random-effects estimates, but will not affect fixed-part estimates (Maas and Hox, 2004; McCulloch and Neuhaus, 2011). Random effects are only biased to a significant degree in extreme scenarios and even in these cases the ranked order of estimated random effects remain highly correlated with rankings of true random effects (Arpino and Varriale, 2010). These small effects and consistency in rankings suggest that substantive inference is unlikely to be affected more than minimally.

Given that this chapter is concerned with relative position in the dataset, the full range of the data needs to be exploited. In order to make this more readily interpretable, the absolute range of the standardised variables is linearly re-scaled to between 0 and 10. Given the skew in some of the responses, this means that the unit differences in response are not exactly comparable across responses. This is a result of the difference in scaling in terms of standard deviation units, as these standard deviations do not follow the distribution of a truly Gaussian response. This matters little in this investigation, however, as the chief concern is the *relative* change in the population for individuals over the 6 waves. A higher score does indicate a higher position on the ranking of each construct. Consequently, modelled differences are longitudinal changes in mental health (as captured by each construct) relative to the population means over the 6 waves.

The final distribution of these responses is given below in Figure 5.1. There are considerable differences between the distribution of the constructs, notably Social Dysfunction and Emotional Coping, which have more normally distributed scores than Lowered Self-Worth and Stress. Lowered Self-Worth exhibits a strong floor effect due to respondents reporting the lowest response category for all of the constituent items. Short of adding a normally distributed random term to each observation[30], which alters the information given in each response, there is little that can be done about this distribution, but this is of little concern given the aim of characterising change in relative position over the observation-period. Given that the total range of the constructs is exactly 10 by design, coefficients will be comparable, but should be interpreted with caution given their different distributions.

[30] This is commonly referred to as "jittering" data in order to overcome the problem of "overplotting" – see e.g. Chambers *et al.*, 1983

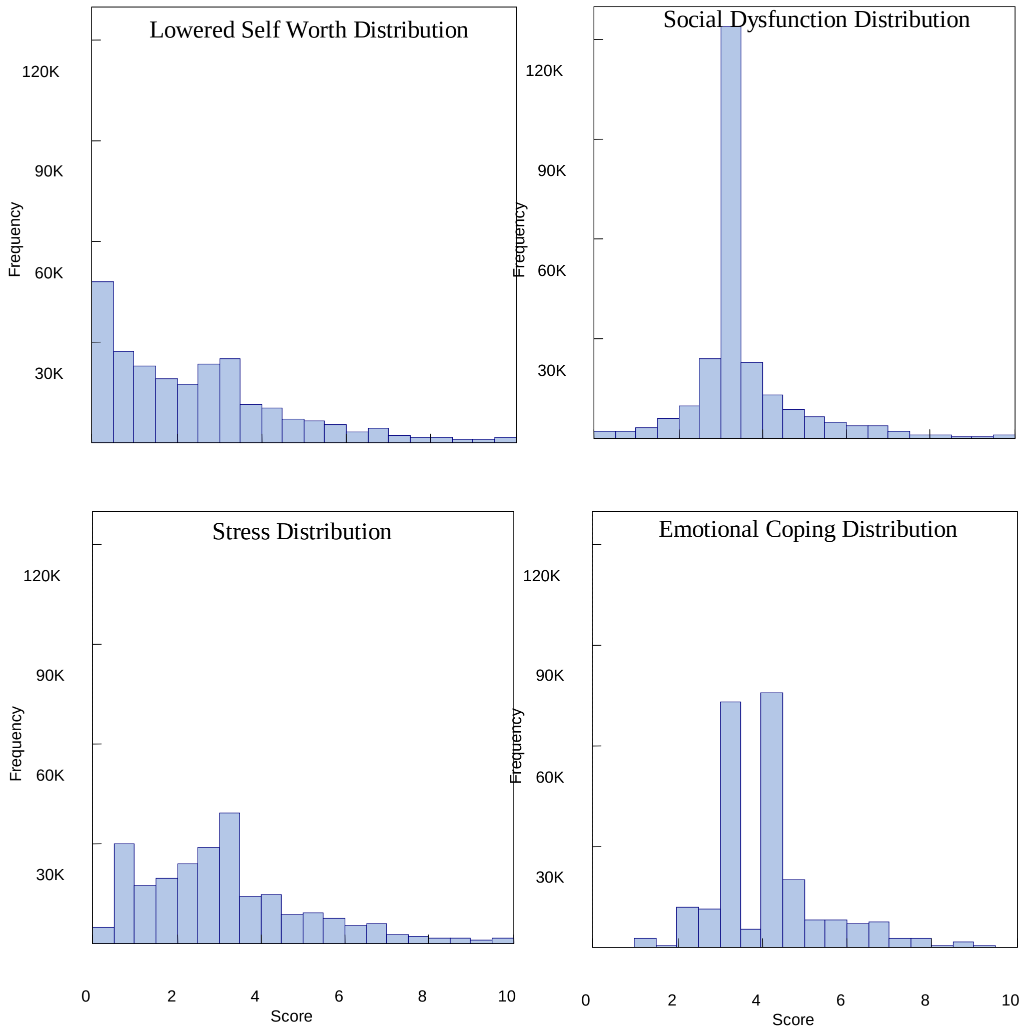


Figure 5.1: Statistical Distribution of Modelled GHQ-12 Factor Responses Constructed for this Chapter.

5.3 Methodology

The methodology here develops on Chapter 4 – now incorporating a structural level for repeated observations within individuals. The model is again explicitly multivariate, simultaneously modelling the responses to the four constructs underpinning the GHQ-12 at each structural level.

5.3.1 The Multilevel Approach

The underpinning model is a 3-Level, multivariate, hierarchical model. Predictor coefficients are interpreted as the change in score on each of the newly constructed metrics associated with a unit increase in the predictor variable (categorical variables are specified as a series of dummies where a unit change denotes the total effect of belonging to that category relative to a reference category specified in Table 5.4). For computational reasons, given the expansive data set, the initial series of exploratory models were fitted using Maximum Likelihood Estimation (Iterative Generalised Least Squares - Goldstein, 2011) in the multilevel package MLwiN (Rasbash et al. 2012).

Multilevel modelling of longitudinal outcomes in this way offers several appealing characteristics for this study. Firstly, it allows the decomposition of unexplained variance into structural levels. Here the structural levels are PSU, Person, and Person-Year. This allows estimation of the relative scale and importance of area, individual and within-individual variation. As in Chapter 4, this is formalised via a Variance Partitioning Coefficient (VPC) however the interpretation will be slightly different due to the different structural specification. For instance, the VPC between level 1 and 2 will no longer be the relative importance of depressive individuals and depressive households, but the relative importance of depressed occasions and depressive individuals.

Secondly, as in the previous chapter, the data used here are explicitly geographically clustered. This is of substantive interest as much of the literature surrounding mental health has identified area effects. Furthermore, repeated measures within an individual are more likely to be similar than repeated measures across individuals (Steele, 2008). This means that there is a conceptual clustering of observations within individuals giving an explicitly hierarchical conceptual framework which can be exploited by multilevel modelling (K Jones, Johnston and Pattie, 1992). Multilevel modelling can account for this inter-dependency of observations. If this dependence is not accounted for, the model risks misattributing findings, potentially overstating the effect of predictors, simply because they are not truly independent observations, but rather measures of the same individual at different time points (Jones, 1991).

Finally, there are a number of individuals who are only captured once, or who drop out of the study to come back at a later wave. Mental health for these individuals is likely to be less reliably estimated than others which are repeatedly sampled. This random-effects, multilevel modelling approach that is used for the analysis has two important and useful characteristics regarding this missingness and imbalance.

Firstly, the multilevel approach allows for missingness in the response variables, without biasing the estimates of the effects of explanatory variables. It does this using the Missing at Random (MAR) assumption, which is a less demanding form of the Missing Completely at Random (MCAR) assumption (Rubin, 1976). The MAR assumption assumes that the missingness is random having conditioned on the included predictors. Thus, if young, lower educated males have a tendency to drop out of the study over time, then including age, sex and education makes the assumption plausible, and estimates unbiased (Bryman, 1988) .This approach allows the analysis of individuals over time even if they have not completed a questionnaires on each and every occasion, i.e. the data are unbalanced. Contextually here, the

MAR assumption holds that those who are not responding in each wave are not doing so because of their mental health.

Secondly given that the object of interest is trends across individuals, estimation needs to be robust, making sure that less-reliably-estimated responses do not overly influence results. Thus, a mechanism is needed by which the reliability of more robustly estimated elements can be incorporated into the less robustly estimated elements, without simply having to drop the less-reliable data points from the investigation. Multilevel modelling uses Precision Weighted Estimation, meaning that those individuals or areas for which there is more data, are used to generate a mean to which the less-robustly-estimated observations gravitate in the model (Jones and Moon, 1991). This random-effects approach allows the dropping of missing values on specific occasions without biasing the model, as the time trend is estimated using borrowed strength from reliably estimated subjects with similar characteristics (Goldstein and de Stavola, 2010).

5.3.2 Model Specification

The baseline specification of this multilevel, multivariate model without predictor variables is given algebraically below[31]:

$$Y_{1jkl} = \beta_1 + (f_{1l} + v_{1kl} + u_{1jkl})$$

$$Y_{2jkl} = \beta_2 + (f_{2l} + v_{2kl} + u_{2jkl})$$

$$Y_{3jkl} = \beta_3 + (f_{3l} + v_{3kl} + u_{3jkl})$$

$$Y_{4jkl} = \beta_4 + (f_{4l} + v_{4kl} + u_{4jkl})$$

$$\begin{bmatrix} f_{1l} \\ f_{2l} \\ f_{3l} \\ f_{4l} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma^2_{f1} & & & \\ \sigma_{f1f2} & \sigma^2_{f2} & & \\ \sigma_{f1f3} & \sigma_{f2f3} & \sigma^2_{f3} & \\ \sigma_{f1f4} & \sigma_{f2f4} & \sigma_{f3f4} & \sigma^2_{f4} \end{matrix}\right)$$

$$\begin{bmatrix} v_{1kl} \\ v_{2kl} \\ v_{3kl} \\ v_{4kl} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma^2_{v1} & & & \\ \sigma_{v1v2} & \sigma^2_{v2} & & \\ \sigma_{v1v3} & \sigma_{v2v3} & \sigma^2_{v3} & \\ \sigma_{v1v4} & \sigma_{v2v4} & \sigma_{v3v4} & \sigma^2_{v4} \end{matrix}\right)$$

$$\begin{bmatrix} u_{1jkl} \\ u_{2jkl} \\ u_{3jkl} \\ u_{4jkl} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma^2_{u1} & & & \\ \sigma_{u1u2} & \sigma^2_{u2} & & \\ \sigma_{u1u3} & \sigma_{u2u3} & \sigma^2_{u3} & \\ \sigma_{u1u4} & \sigma_{u2u4} & \sigma_{u3u4} & \sigma^2_{u4} \end{matrix}\right)$$

[31] As in the previous chapter, there is no subscript i denoting the lowest structural level as this level exists solely to define the multivariate structure, again denoting which of the four constructs an observation constitutes. Whilst this technically makes the specification a 4-level structure, it is functionally identical to a 3-level specification.

Term	Interpretation
$\boldsymbol{Y_{1jkl}} \dots \boldsymbol{Y_{4jkl}}$	Observed response for Lowered Self Worth (1), Social Dysfunction (2), Stress (3) and Emotional Coping (4) for occasion *j*, from individual *k*, within PSU $\boldsymbol{l}$;
$\boldsymbol{\beta_1} \dots \boldsymbol{\beta_4}$	The mean score across responses (1-4) for a typical observation, in a typical individual in a typical PSU.
$\boldsymbol{f_{1l}} \dots \boldsymbol{f_{4l}}$	The differential residuals from the mean responses for each PSU. These are assumed to come from a Normal distribution, centred on the associated mean value denoted by the $\boldsymbol{\beta}$ coefficients.
$\boldsymbol{v_{1kl}} \dots \boldsymbol{v_{4kl}}$	The differentials for each response, for an individual *k given* their PSU $\boldsymbol{l}$ across occasions. These are given as differentials from the PSU mean. These are again assumed to come from a Normal distribution, centred on the combined mean *differential* value denoted by the $\boldsymbol{\beta}$ coefficients and the $\boldsymbol{f}$ PSU differential.
$\boldsymbol{u_{1jkl}} \dots \boldsymbol{u_{4jkl}}$	The differentials across responses for occasion *j*, given as differences from the combined values taken from the differentials of individual *k* in PSU $\boldsymbol{l}$ from the grand mean $\boldsymbol{\beta}$. These are assumed to be normally distributed, centred on the aforementioned person-specific, PSU-specific modelled differentials.
$\boldsymbol{\sigma^2_{f1}} \dots \boldsymbol{\sigma^2_{f4}}$	The variance of the residuals at the PSU-level for all individuals across all occasions.
$\boldsymbol{\sigma^2_{v1}} \dots \boldsymbol{\sigma^2_{v4}}$	The variance of the residuals at the Individual-level for all observations within that individual, *given* their situation within a specific PSU.
$\boldsymbol{\sigma^2_{u1}} \dots \boldsymbol{\sigma^2_{u4}}$	The variance of the residuals at the observation-level for all responses within the observation, given the respondent, and their respective PSU.
$\boldsymbol{\sigma_{fXfY}}$	This gives the covariance of the residuals which have been aggregated to the PSU-level $\boldsymbol{l}$ between response 'X' and response 'Y'.
$\boldsymbol{\sigma_{vXvY}}$	This gives the covariance of the residuals at the individual level *k*, having taken account of their commonality due to belonging to PSU $\boldsymbol{l}$, between response 'X' and response 'Y'.
$\boldsymbol{\sigma_{uXuY}}$	This gives the covariance of the occasion-level residuals, having taken account of the individual *k* and the PSU $\boldsymbol{l}$, between response 'X' and response 'Y'.

Table 5.5: Interpretation of Model Coefficients from baseline 3-level longitudinal, multivariate specification.

Table 5.5 gives the interpretation of the final model parameters. Of particular interest are the overall means β_x, the variances between individuals[32] σ^2_{vx}, and the variances between-occasions within-individuals σ^2_{ux}, as the ratio of these will inform to what extent these constructs are static within a person, or variable over time. Similarly, the modelled covariances are of interest as they are informative of the extent to which these responses are predictive of

[32] The level referred to as "individual" here is actually (relative to the previous chapter) analogous to "individual and household level" due to the non-inclusion of household (Tranmer and Steel, 2001). These are not differentiated in the modelling procedure due to the large scale of the modelling process and the complexity of constructing a complex and varying household level across the 6 waves.

each other at both the individual σ_{vXvY} and observation level σ_{uXuY}. It would be expected, given the low degree of commonality seen in Chapter 2, that correlations at these lower structural levels are lower than 0.9. This would validate the treatment of these constructs as separate but related, rather than homogenous as implied by the typical unidimensional assumption.

The specific variance terms for each construct at each structural level give a directly comparable measure of the unexplained variation present at the area, individual and within-individual occasion levels. Given that the first research question specifically pertains to the degree of consistency within individuals vs. changes within an individual over time, a Variance Partitioning Coefficient (VPC) will be constructed for each of the structural levels, for each response

$$\text{Occasion Level VPC for } Y_{1jkl} = \frac{\sigma_{u1}^2}{\sigma_{u1}^2 + \sigma_{v1}^2 + \sigma_{f1}^2}$$

The VPC gives the unexplained variation present at that structural level as a proportion of total variance and is typically presented as a percentage value (Goldstein, 2011). It can alternatively be interpreted as the degree of intra-class correlation, that is the correlation within a higher-level unit if observations were taken at random from within it (Browne et al. 2005). More simply, if hypothetical pairs of observations from within an individual on different occasions were compared, the correlation between the long-stacked vectors of those observations would be the same as the VPC. In this capacity it gives the degree of consistency within a higher-level unit. If the largest VPC for a measure is at the individual level, then this would imply that between-individual variation is the largest source of unexplained variation. Similarly, if the largest VPC is at the occasion level then this would indicate that most of the variation is within individuals, thus that individuals fluctuate in and out of what would be categorised as "ill" between years.

The VPC is potentially of more immediately obvious interest than the variances, yet the absolute variance values also offer substantive insight. The response variables have been standardised as outlined above, there are differences in the distribution of these standardised variables which correspond to genuine differences in individual response patterning. Therefore, the total variance within each construct is also of substantive interest due to it reflecting true differences in the variability of possible scores as shown in Figure 5.1. Larger variances indicate a distribution that is potentially more platykurtic, having a less strong clustering around the mean. This would indicate that the mental health experience characterised by this response would be more variable overall. Thus, absolute variances, relative structural distribution of unexplained variance, and covariance between residuals of each response at each structural level are presented and discussed later in this chapter.

5.3.3 Model Building Strategy

The model-building strategy is exploratory following the process set out by Jones et al. (2000). It investigates the effect of the inclusion of a series of variables, and their interactions, on the fit of the model specified. Absolute deviance and, for nested models, the change in deviance between the model and the simpler specification are reported. This change in deviance is then evaluated against a Chi-Square distribution to find the probability of the reduction in deviance being by chance. The criterion for continued inclusion in the model was a Chi-square probability of the reduction in deviance being significant at the 0.05 level. However, given the size of the data-set, it is likely that nearly all changes will be significant at this level.

Having established how improvement in fit will be evidenced, the model building process needs to be outlined. The increasing complexity in the model is broken down into several components. Firstly, the specification from a simple single-level model is developed into a model with repeated observations nested within individuals, and within sampling units, estimating only the unconditional, unexplained variance remaining at each level in the variance/covariance structure. The next stage involves adding more complexity via the specification of complex variance structures at each of the structural levels. This involves introducing and estimating the covariance terms between each of the responses at each of the structural levels.

Having established the complexity within the null model, the individual-level predictors indicated in Table 5.4 are then included. Firstly, each of the predictors is included in isolation, assessing the change in deviance the inclusion of each of the terms generates from the null model. Knowing the order in which these predictors explain the most variation as evidenced by this deviance, the null model is re-specified, and predictors are introduced sequentially in order of which had the greatest explanatory capacity in isolation. This gives the greatest likelihood of attributing variance correctly. The structural year predictor is then introduced, to model change over the years, and ensure that coefficients for demographic characteristics are estimated net of year on year changes. Finally, a number of interaction terms are specified, in order to assess differential longitudinal trends in mental health for different demographic groups.

5.3.4 Model Estimation

All models presented here were estimated using the Iterative Generalised Least Squares estimator implemented in MLwiN v3.00. Changes in the log-likelihood (Deviance) are presented to evidence improvement of fit (Self and Liang, 1987). The Bayesian framework which has been adopted in the previous research chapters was not possible for all models specified here due to the computational intensity of the model specification coupled with the scale of the data. However, a Bayesian MCMC run was undertaken for the final model including all predictors so all estimates for all terms take account of the uncertainty in the other estimates.

5.3.5 Research Questions.

The four broad research aims are once again presented below, but now with an explicit outline of how they will be answered methodologically:

1. Over this 6-year period, how temporally stable are these mental health constructs between areas, between individuals and within individuals? Are individuals consistent in displaying mental distress?
    a. Conduct an exploratory analysis, determining the best explanatory model justified by reduction in deviance.
    b. Specify complex variation between each of the simultaneously modelled responses at each structural level.
    c. When allowing for a random structure, covarying between each of the four measures at each of three structural levels, calculate the proportion of absolute variance in each response that remains unexplained at the PSU-level, individual-level, and the within-individual, between-occasion level?
    d. Is this proportion of variance at the individual/occasion level different for different responses, and if so how?

2. How correlated are each of the four mental health constructs, at each structural level?
    a. Having specified complex covariance between each of the constructs at each structural level, how correlated are they between PSUs, between individuals taking account of location, and between occasion taking account of individual differences across-occasions?

3. Averaging over the 6-year period, which demographic groups are most vulnerable to mental distress? Do these at-risk demographics differ across each of the four constructs developed in Chapter 2?
    a. Specifying longitudinal models, and controlling for the year of observation, are there still differences between demographic groups based on age, sex, ethnicity, education, employment and marital status? Do these relationships differ depending on which constructs within the GHQ-12 are considered?

4. Having looked at aggregate differences over the entire time period, are different demographic groups experiencing different trajectories in mental health in the period 2009-2014?
    a. Introduce a first-order interaction term between categorical "waveyear" and each demographic variable, allowing for separate estimation of differential effects for each demographic group for each response.
    b. Are there distinct differences (as determined by non-overlapping confidence intervals for a given differential) in mental health trajectories between different demographic groups, and do these highlight any specific risk populations?

5.4 Results

The results of the exploratory model building procedure are displayed below in Table 5.6. The change in the deviance measures the overall improvement in the badness of fit between the observed values of the responses and the fitted values for a given model in comparison to a simpler specification. A reduction in this deviance can be achieved both by including specific predictor variables in the fixed part, and by specifying more complex (hence more realistic) random parts which structure the unexplained variation in the outcomes.

The initial baseline model has four fixed terms, one average for each of the responses. It similarly has four random part terms at each level, consisting of an estimated variance for each response at each structural level. Block 1 in Table 5.6 involves the introduction of complex variation to this structure, allowing residuals to be correlated across responses at each structural level. Block 2 involves the introduction of individual predictor variables in isolation and in an arbitrary order. Block 3 then includes those same predictors sequentially but in the order of those that produced they greatest reduction in deviance when included individually. Block 4 then includes higher-level structural variables regarding the wave of measurement. The final stage of the modelling is not displayed in Table 5.6 but involves specifying interaction terms between each predictor variable and the categorical wave term.

5.4.1 Exploratory Analysis

Model	Model Specification	Model Comparator	Deviance	Reduction in Deviance from Comparator	Δ d.f.	Significance (p-value)
Structural Complexity						
1.0	Baseline, Diagonal Variances Only	-	3373171.047			
1.01	1.0+Lv1Complex Variance	1.0	2711479.275	661691.772	6	<0.001
1.02	1.01+Lv2Complex Variance	1.01	2636180.230	75299.045	6	<0.001
1.03	1.02+Lv3Complex Var - 3 Level Model, with full Variance Specification	1.02	2635681.945	498.285	6	<0.001
Individual Predictors						
2.2	1.03+Sex	1.03	2634238.984	1442.961	4	<0.001
2.3	1.03+MarStat	1.03	2627765.359	6473.325	8	<0.001
2.4	1.03+HiQual	1.03	2604906.658	29332.326	16	<0.001
2.5	1.03+Age	1.03	2628452.980	5696.004	4	<0.001
2.51	1.5+Age^2	2.5	2626609.705	1843.275	4	<0.001
2.52	1.51+Age^3	2.51	2626218.044	391.661	4	<0.001
2.6	1.03+Ethnic	1.03	2605531.581	28707.403	12	<0.001
2.7	1.03+AgeCat	1.03	2626782.749	7456.235	28	<0.001
2.8	1.03+JType	1.03	2616435.546	17803.438	24	<0.001
Combined Individual Predictors						
3.1	1.4+Ethnic	2.4	2582670.051	22236.607	12	<0.001
3.2	2.1+JType	3.1	2565746.345	16923.706	24	<0.001
3.3	2.2+MarStat	3.2	2559345.686	6400.659	8	<0.001
3.4	2.3+Age	3.3	2556104.494	3241.192	4	<0.001
3.41	2.4+Age^2	3.4	2554002.727	2101.767	4	<0.001
3.42	2.41+Age^3	3.41	2553218.130	784.597	4	<0.001
3.5	2.42+Sex	3.42	2552188.644	1029.486	4	<0.001
Structural Predictors						
4.01	1.03+WaveYear	1.03	2634264.128	1417.817	20	<0.001
4.1	3.5+WaveYear	3.5	2551001.428	1187.216	20	<0.001

Table 5.6: Exploratory Model Building Process, illustrating sequential process of model fitting and evaluation of improvement in fit at each iteration.

As the results show, the numerical power associated with a sample of this size is such that, as expected, each specification of additional complexity was significant when evaluated against a chi-square distribution. However, the raw difference in deviance values displayed give indication as to which variables or specifications account for the most variation. Structurally the largest improvement in fit across any of the alterations was the introduction of complex variation at level 1. Operationally this involves the specification of covariance terms, modelling inter-relatedness and giving the degree of similarity between each of the four responses, at a given structural level, rather than solely estimating variances in each as unrelated. Socio-economic characteristics clearly explain the greatest amount of difference in mental health outcomes. Highest educational attainment is the single predictor that explained the largest amount of difference across the four responses. Ethnicity explained the next largest amount, followed by Job Classification, Marital status, Age and Sex respectively. Despite initial concerns about the collinearity due to socioeconomic patterning, when introduced into the model sequentially education, ethnicity and job classification all continued to explain by far the largest amount of unexplained difference in responses. Whilst the model building procedure demonstrates a clear hierarchy of predictive capacity, the model with the greatest predictive capacity was model 4.1 which specifies terms for complex variation at all structural levels, all demographic predictors and a categorical wave term. This was the model re-estimated by MCMC.

Having established that this model is the best fit, the first investigative research question asks to what degree these mental health constructs are consistent across individuals. The following section analyses the predictive capacity of the model by graphically representing the coefficients for each variable, holding the effect of the remaining variables at their mean.

5.4.2 Temporal and Geographical Consistency of Mental Health Experience

The correlations between responses at each structural level are given in Table 5.7. Values along the leading diagonal of each structural matrix give variances for each response at that level. Additionally, the lower portion of the table presents the VPCs for each structural level for each response variable.

	LOWERED SELF-WORTH	***SOCIAL DYSFUNCTION***	***STRESS***	***EMOTIONAL COPING***
LEVEL 3: PSU				
LSW	**0.083**			
SOCDYS	*0.927*	**0.015**		
STRESS	*0.986*	*0.941*	**0.071**	
EMOCOP	*0.961*	*0.955*	*0.959*	**0.018**
LEVEL 2: INDIVIDUAL				
LSW	**2.102**			
SOCDYS	*0.781*	**0.513**		
STRESS	*0.961*	*0.774*	**1.655**	
EMOCOP	*0.888*	*0.837*	*0.887*	**0.425**
LEVEL 1: OCCASION				
LSW	**1.944**			
SOCDYS	*0.595*	**0.984**		
STRESS	*0.881*	*0.620*	**1.868**	
EMOCOP	*0.644*	*0.595*	*0.618*	**1.027**
VPC				
LEVEL 4: PSU	2.01%	0.99%	1.98%	1.22%
LEVEL 3: INDIVIDUAL	50.91%	33.93%	46.05%	28.91%
LEVEL 2: OBSERVATION	47.08%	65.08%	51.98%	69.86%

Table 5.7: Decomposition of Variance from Model 4.1,
Upper portion gives structural similarity between constructs with raw variances emboldened, and correlations at each structural level italicised. Lower portion gives level-specific VPCs for each response.

Table 5.7 shows that PSU-level variances are low across all four metrics, however are higher for Lowered-Self Worth and Stress than for Social Dysfunction and Emotional Coping. This pattern is replicated across the lower scales, with the Social Dysfunction and Emotional Coping constructs having lower absolute variance values across all three levels. This is due to the different distributions of the responses seen in Figure 5.1. Given this apparent similarity, the relationships between the four responses vary considerably by structural level, which is evidenced by the correlations.

There is evidence of considerable dissimilarity between the constructs at both the occasion- and individual-levels. The modelled correlations at the occasion-level are greater than those which were originally modelled in Chapter 2, as these are no longer being modelled as latent underpinning variables, but as calculated, manifest constructs. The correlations are still not large, when considering the squared correlation as indicative of the true degree of predictive capacity, as outlined in Chapter 2. Interestingly, the largest correlation is not between Social Dysfunction and Lowered Self-Worth as in Chapter 2, but between Lowered Self-Worth and Stress, due at least partly to them being composed of several of the same items. There are some interesting differences in the patterning within and between individuals. At Level 2, there is a larger correlation between Social Dysfunction and Stress than between Social Dysfunction and Emotional Coping, suggesting that at a time point when an individual is experiencing high levels of Social Dysfunction, they are more likely to also be experiencing high levels of stress than high levels of emotional coping. Compare this with Level 3, where the correlation between Social Dysfunction and Emotional Coping is higher. This indicates that if an *individual* struggles with social functioning in general then they are more likely to be more stoic about their negative emotion. However, if there is an *occasion* where an individual is finding social functioning more difficult than normal, this is more likely to be associated with a period of high stress.

There is also considerable dissimilarity between constructs when considering the relative importance of PSU, individual and occasion[33]. VPCs for level 3 may initially seem very small, at roughly 2% for Lowered Self-Worth and Stress, and closer to 1% for Social Dysfunction and Emotional Coping. These are, however, considerable consistencies, given that these effects are estimated having taken account of the individuals *and* the time period of the observation. Having 2% of the total variation in individual self-worth over a 6-wave time period being consistent at the PSU-level suggests a non-negligible degree of geographical patterning to the results. Correlations between measures at level-3 are higher than at the other levels, over 0.9 for all responses, meaning positive PSU-level residuals on any one of the constructs are associated with positive residuals on the others. This suggests that any PSU with a high level of one of the constructs is highly likely to have a high level of another. Despite this there is still evidence that the area an individual is located in is approximately twice as important for Lowered Self-Worth and Stress than Social Dysfunction and Emotional Coping.

The most striking differences between the constructs are in the Level 2 and 3 VPCs. Whilst there is clear evidence of greater levels of within-individual variance for Lowered Self-Worth and for Stress, when compared with the between-individual variation, the occasion becomes less important. Social Dysfunction and Emotional Coping are much more variable within an individual than Stress and Lowered Self-Worth. This implies that for these responses the specific occasion is more important than the individual, which ties in with the change in correlation between the two levels for these responses. Conversely, there is evidence that there are stressed individuals and individuals with low self-worth, who stay more consistently so

[33] Individual predictors are included in this model, but their inclusion changed little in the structuring of the variance and correlations. Additionally, the difference between the VPCs before and after the introduction of individual predictors was trivial.

over time. The constructs referring to stress and self-worth are more concerned with internal self-evaluation, rather than external evaluation of functioning. It is perhaps for this reason that they are more internally consistent within an individual. Differences between proportions are even more stark when considered relatively. The occasion is twice as important as the individual for Social Dysfunction and for Emotional Coping. Conversely the individual is equally as important as the occasion for Lowered Self-Worth and Stress.

There is a clear role of geography in the variability of mental health as captured by these constructs. However, the most notable result here is the relative importance of the individual. Individual-Level VPCs are high across all metrics, yet for all but Lowered Self-Worth the importance of the individual is outweighed by the importance of the occasion of measurement. Whilst this longitudinal finding is limited to the relatively short period of 6 waves (8 years) across a lifetime, this suggests that even though individuals are consistent, the occasion is the most important factor for all but Lowered Self-Worth. This implies that perhaps notions of "mentally unhealthy individuals" should be reframed as "mentally unhealthy experiences", especially when considering aspects of social functioning, and perceived competence in light of negative mental experience.

5.4.3 Longitudinal Demographic Patterning of Mental Health

Having established the varying degree of chronicity over different dimensions of mental health, the next research question investigates the consistency of the findings of the cross-sectional investigation in Chapter 4 when analysed longitudinally. This enables the characterisation of relationships between demographic groups as more than simply cross-sectional patterning in a given year. This section details the results of the estimated effects of each demographic predictor, holding all other predictors constant at their mean value. Results are presented sequentially for the following predictors: Sex, Age, Marital Status, Education, Ethnicity, Employment Classification and Year. For ease of interpretation, Lowered Self Worth, Social Dysfunction and Emotional Coping are abbreviated to LSW, SD and EC respectively for the remainder of the results.

5.4.3.1 Sex Patterning

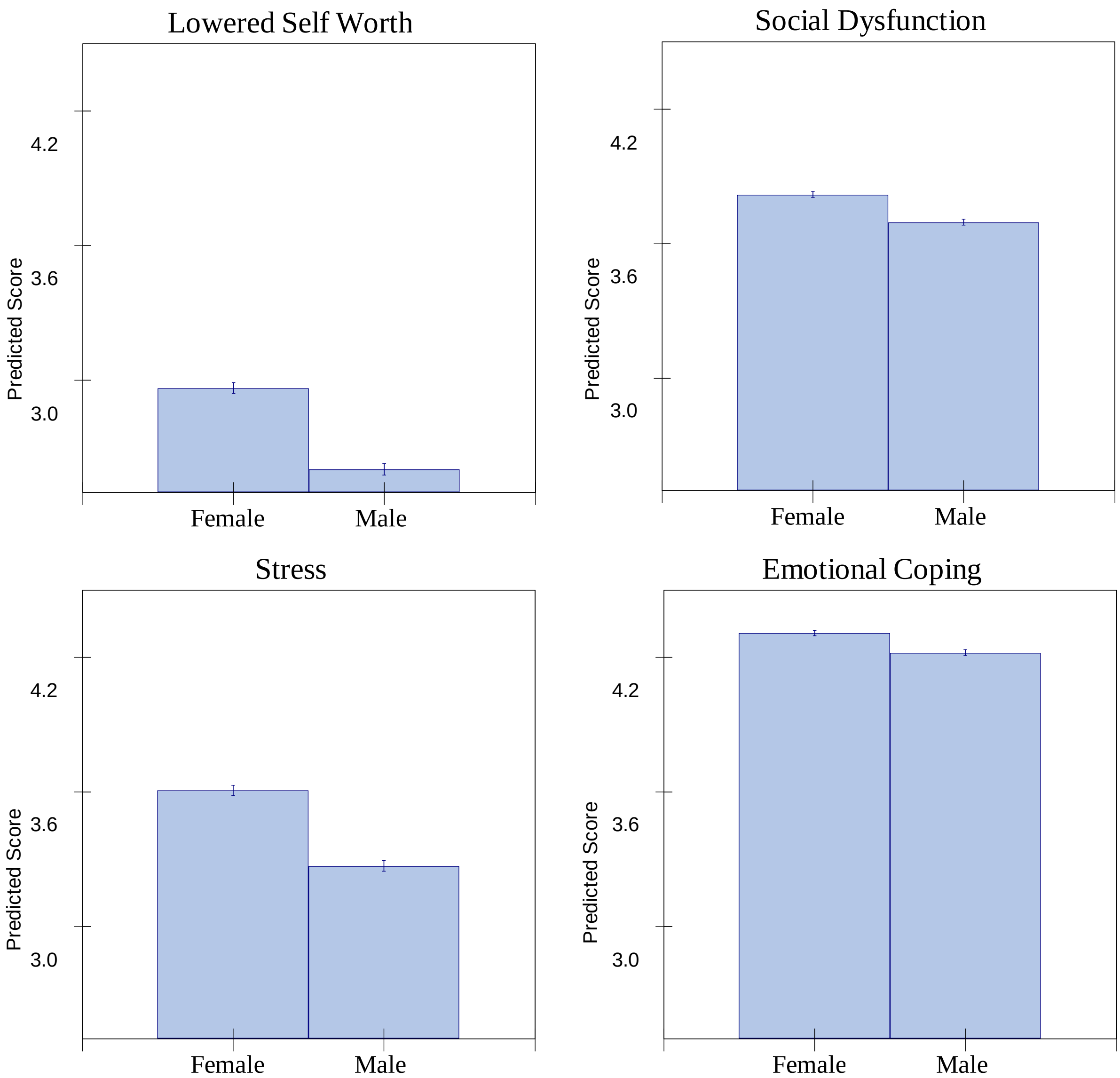


Figure 5.2: Predicted values for differences in Sex across the four modelled response variables.

Figure 5.2 illustrates the modelled differences between males and females for the different dimensions of the GHQ-12. Females are shown to have worse mental health for each dimension, consistent with numerous prior studies showing higher rates of depressive disorders as captured by the GHQ-12 for females across various populations (e.g. Weich et al. 1998, McLean et al. 2011, Ferrari et al. 2013).

The largest differences between sexes are present for the LSW and Stress dimensions, with females scoring 0.3 points higher across these metrics. SD and EC have smaller differences between sexes. The composition of these metrics may play a role here. As identified in Chapter 4, the strong between-sex difference that is present in the raw GHQ-12 scores, is not present in the SWE responses. In Chapter 4 this was argued to be a function of the phrasing of the questions whereby males were more likely to respond negatively to a positive question than respond affirmatively to a negative one (Murray et al. 2008). The result here demonstrates that the GHQ-12 difference between sexes demonstrated in Chapter 4 is consistent over time. The LSW dimension, which is comprised entirely of the negatively phrased items, has a far greater between-sex disparity than the SD dimension, which is composed of the positively phrased items. However, the Stress dimension is comprised of 4 negative and 2 positively phrased items, and still demonstrates a comparable difference between sexes to that of the LSW dimension. This suggests that whilst the between-sex difference in response to the phrasing of items may play a part in the overall disparity between sexes, it is clearly only a part of this reason.

EC has the smallest sex difference. Whilst statistically significant, the magnitude of the difference is far smaller than the LSW and Stress dimensions. This small difference similarly cannot be explained simply as a product of the wording of the items, as EC is made up of three negatively phrased items and one positively phrased item. The inverse scoring of this positive item, however, changes the interpretation. The highest scores on EC indicate individuals who are feeling strongly negatively emotional, as captured by three negatively worded items, yet do not feel this negatively affects their decision-making capacity. If the stoicism explanation held true, one would possibly expect to see very low scores for females whose higher scores on the negative responses would be converted to very negative scores on EC, and higher scores for males, which – although the gap is smaller – is not what is shown. It is clear that women in general consistently report worse mental health over all dimensions, although this is much more pronounced in the self-reflective LSW and Stress constructs than the socially evaluated SD and EC constructs.

5.4.3.2 Polynomial Age Patterning

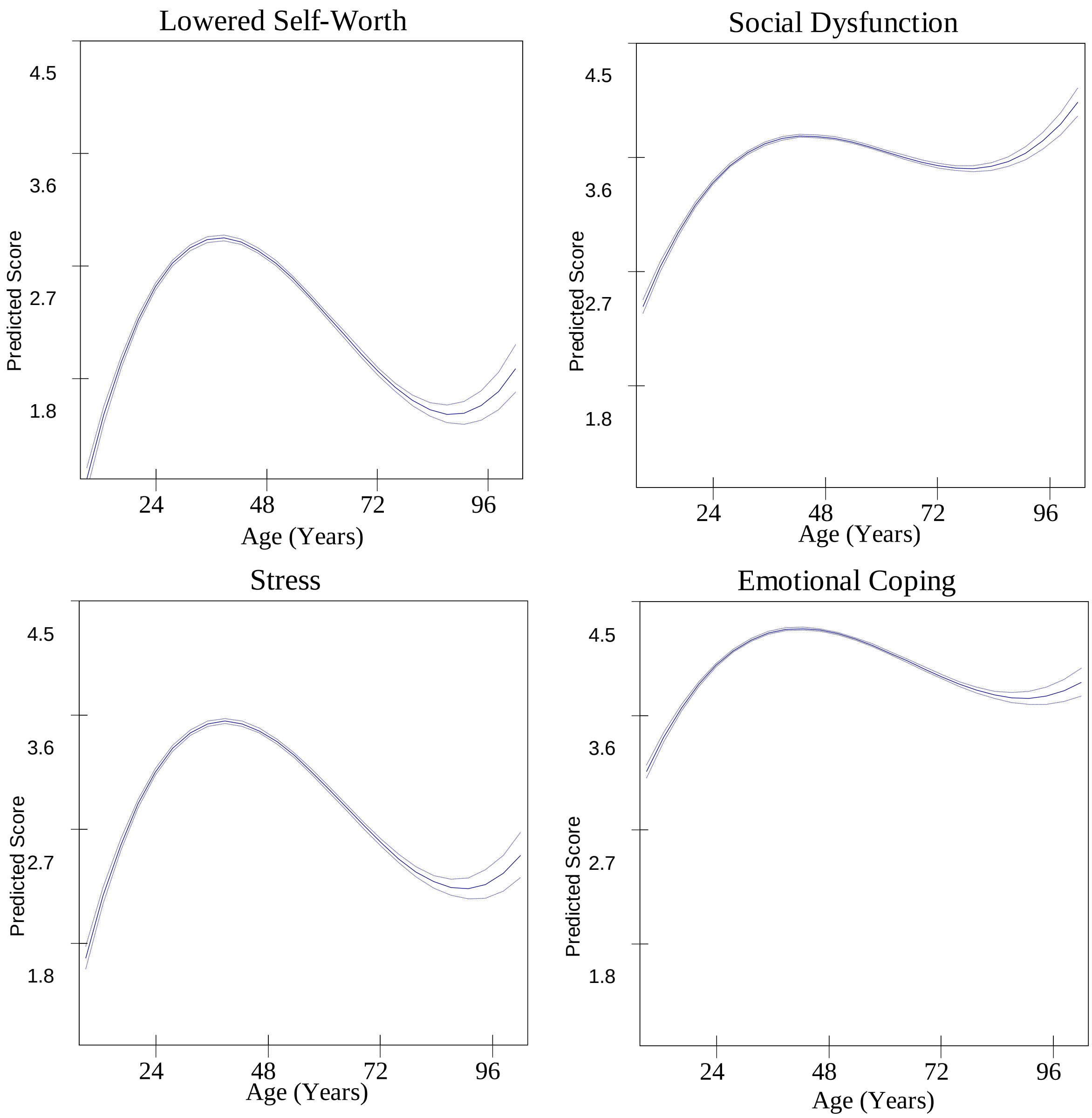


Figure 5.3: Predicted values stratified by Age (in 2009) across the four modelled responses.

Figure 5.3 shows the age profile of each GHQ-12 measure – it does not give a profile of *ageing*, it gives a cross-sectional representation based on the age of the respondent in the first wave, aggregated over the 6-wave time period. Each response has polynomial age predictors specified, up to and including a cubic term. The first and most consistent observation is of the clear decline in mental health with age until the mid-30s, which is consistent across all four responses. There is clear evidence of the mid-life peak in mental distress commonly identified in mental health literature (e.g. Jorm, 2000, Blanchflower et al. 2008, Lang et al. 2011, Ulloa et al. 2013). This midlife-peak in mental distress is generally reported to be in the mid to late 40s, which again is consistent with these results (Lang et al. 2011, Blanchflower et al. 2008).

There are again striking differences between constructs, depending on the dimension of mental health considered. The similarity between the constructs seems to be heavily dependent on age, with the differences between LSW and SD increasing sharply with age past 40. This dissimilarity increases with age, and is greatest in those above 50, where LSW follows a more traditionally recognised bell-shape. Conversely SD seems to simply slow its deterioration in midlife, before accelerating once more past 80, although there is greater uncertainty around the elderly as there are fewer respondents over 80. This similarity of the patterns between the younger respondents seem to suggest that there is less difference between the responses to different items for the younger respondents than for the older respondents. Indeed, LSW and Stress show considerably better mental health amongst the elderly, where SD and EC seem to show the opposite, with a clear second turning point after the first midlife peak.

This is more consistent with results found by Bell (2014), who suggests that evidence for the midlife peak in mental distress is commonly found due to misattribution of Age-Period-Cohort (APC) effects, and if cohorts are controlled, then mental health (as captured by the GHQ-12) deteriorates steadily with age, albeit slowing slightly in middle age. Whilst there is no effort to attribute differences to APC effects here, beyond incorporating a longitudinal structure, this does seem to suggest that perhaps the widely accepted midlife peak in mental distress needs re-evaluation, however the Bell paper treats the heterogeneous elements of the GHQ-12 as unidimensional. Thus, the findings need not necessarily be interpreted as an artefact of methodology, but perhaps viewed as differential patterns for different age groups across different dimensions of mental health. Multidimensional investigations of mental health necessarily imply that age trajectories can differ by response. For instance, Westerhof and Keyes (2010) demonstrated the potential of the two-continua model in discovering that elderly individuals expressed fewer symptoms of mental illness than younger counterparts but displayed no significant difference in positive mental health experience. Contrasted with the results here, SD and EC seem to echo the responses of positive mental health measurement, where LSW and Stress seem to demonstrate the relationship they found for mental illness.

Similar to the sex relationship, this relationship may also be a product of the tendency to respond more honestly to certain questions given their phrasing. The tendency of the elderly to under-report physical pain has been discussed at length, often under the same guise of stoicism, stigma and discomfort with dependence as for male respondents (e.g. Foley, 1994, Helme and Gibson, 2001, Yong, 2006, Pachana, 2008). This discussion has rarely been carried across to mental health although there are some notable exceptions (Mah et al. 2018, Conner et al. 2010, Bryant 2010). If it is considered that differences between ages are possibly due to differences in reporting, rather than true underlying differences in mental distress, then this perception of stoicism would likely produce a similar effect amongst the elderly population to what is seen here. The construct comprising positively phrased items, SD, shows very different trajectories to the negatively phrased LSW construct. Perhaps predictably, the most stressed individuals as captured by the third construct are also those in mid-life. There is a similar trend to SD in the EC response, where this stoicism, if present, would be expected to be manifested. The individuals most likely to feel strongly negative and respond positively to it anyway are those in middle age and the very elderly, although again the total range of this variable is less than the other responses.

5.4.3.3 Marital Status

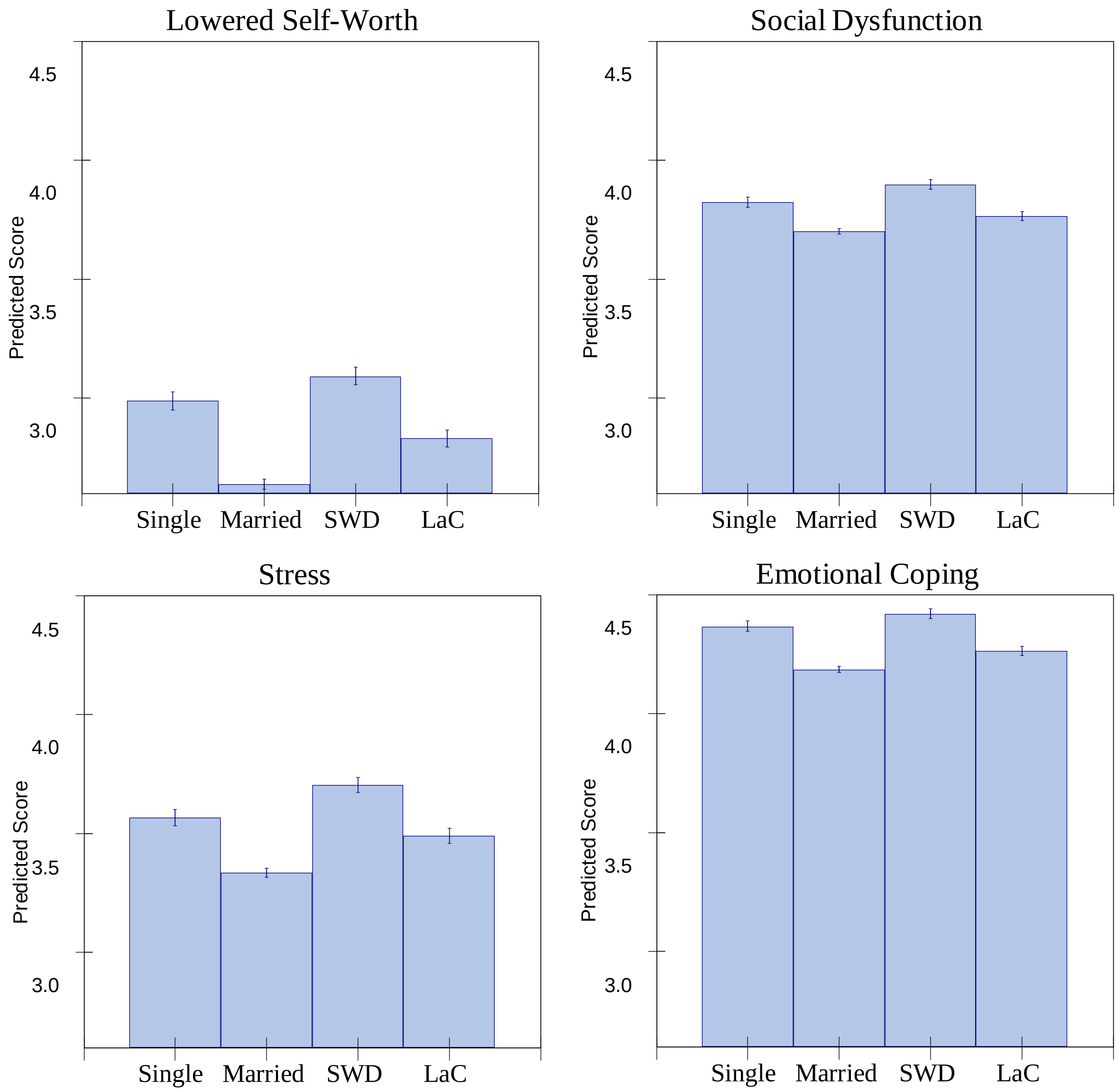


Figure 5.4: Predicted mental health scores across the four constructs for Marital Status. (Single, Married, Separated/Widowed/Divorced, Living as Couple).

Figure 5.4 illustrates the differences in modelled mental health across the four responses for Marital Status. Consistent with previous studies married individuals display the least mental distress across all four constructs (Wade and Pevalin, 2004, Wilson et al. 2005, Lindstrom et al. 2012). Those "Living as a Couple" as separate to "Married" experience better mental health than single or SWD respondents but not quite as good mental health as individuals who are married. The clearest difference between married individuals and those living as a couple is in their LSW and Stress levels. The benefits of marriage are less strongly expressed in SD and

EC, where individuals living as a couple express similarly improved mental health than those who are single or SWD. This suggests that, perhaps obviously, cohabiting and associated companionship and lack of isolation comprise the key beneficial elements of marriage, rather than explicitly marriage itself.

The differences between married and single individuals are greatest for LSW, as with previous results, although the differences are less marked. Consistent with previous research, individuals who are separated, widowed or divorced experience the worst mental health across the first three responses (Richards et al. 1997), and are additionally the least likely to consider their poor mental state as affecting their decision-making capacity. EC is lowest for married individuals, which is likely to largely be due to their lower overall distress, but which may echo the suggestion of security and confidence allowing individuals to acknowledge their psychological distress as potentially socially debilitating.

5.4.3.4 Highest Educational Qualification Patterning

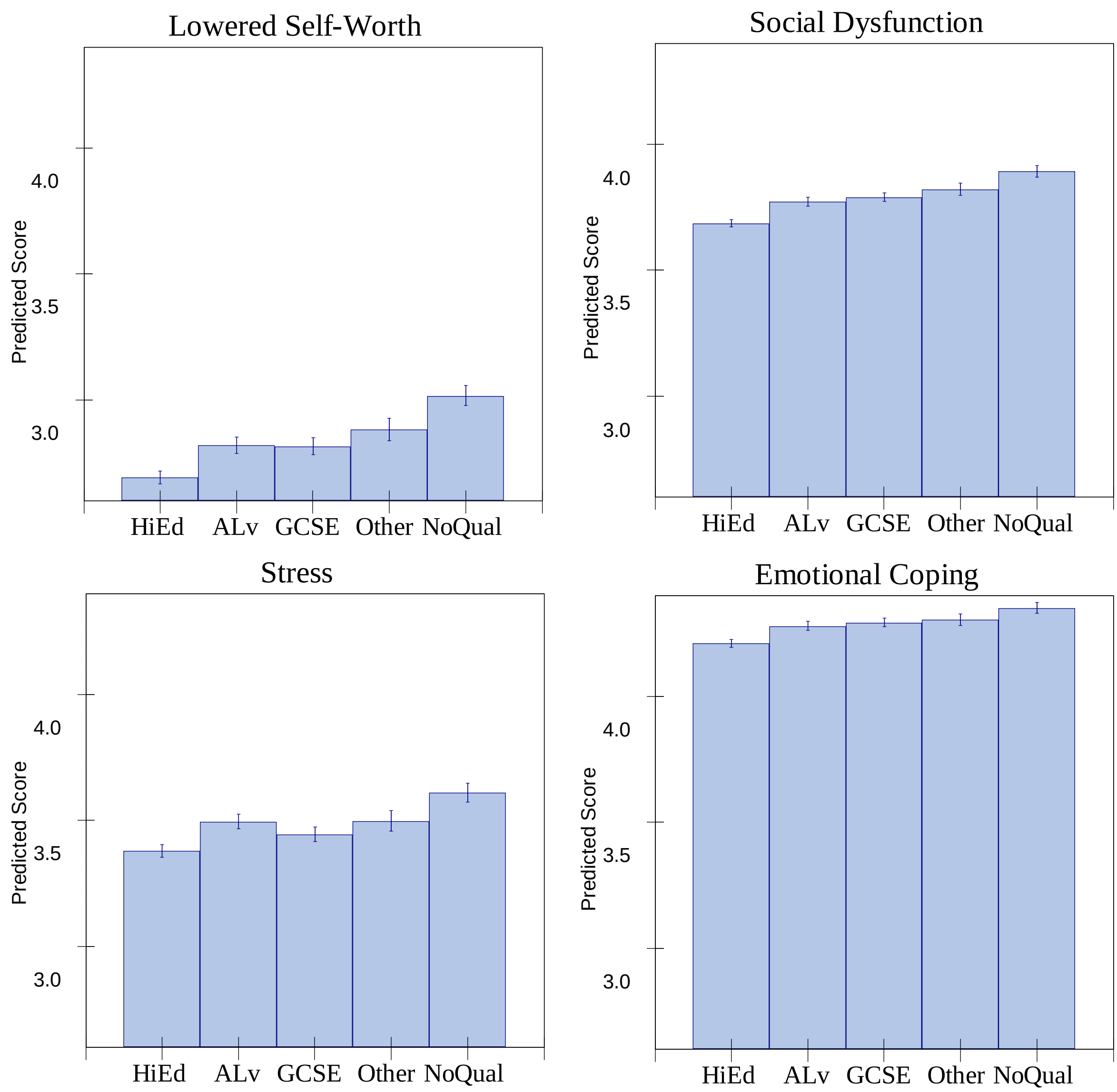


Figure 5.5: Predicted mental health scores across the four constructs for Education as captured by Highest Educational Qualification.
(Higher Education, A-Level, GCSE, Other, No Qualification).

Figure 5.5 illustrates the differences between educational attainment for the four responses. There is evidence of a dose-response relationship with education, in that individuals with higher education seem to experience better mental health than those with A-levels or GCSEs who in turn experience better mental health than those with no qualification. The broad trend is consistent with much of the literature, which tends to suggest that higher education is associated with lower mental distress (Ellen et al. 2001, Murali & Oyebode, 2004) and increased mental well-being (Westerhof & Keyes, 2010, Stewart-Brown et al. 2015).

The relationship across the responses is one of the most consistent reported in this chapter. Across all four responses, individuals with a GCSE qualification have better mental health than those with no qualification, and those with a higher educational qualification demonstrate better mental health than those whose highest attainment is GCSE. The main difference between groups is between those who achieved A-Levels, and those who continued on to get a higher educational qualification. A-Level and GCSE achievers are statistically indistinguishable for all responses bar stress, where those who achieved A-Levels experience worse stress levels.

5.4.3.5 Ethnicity

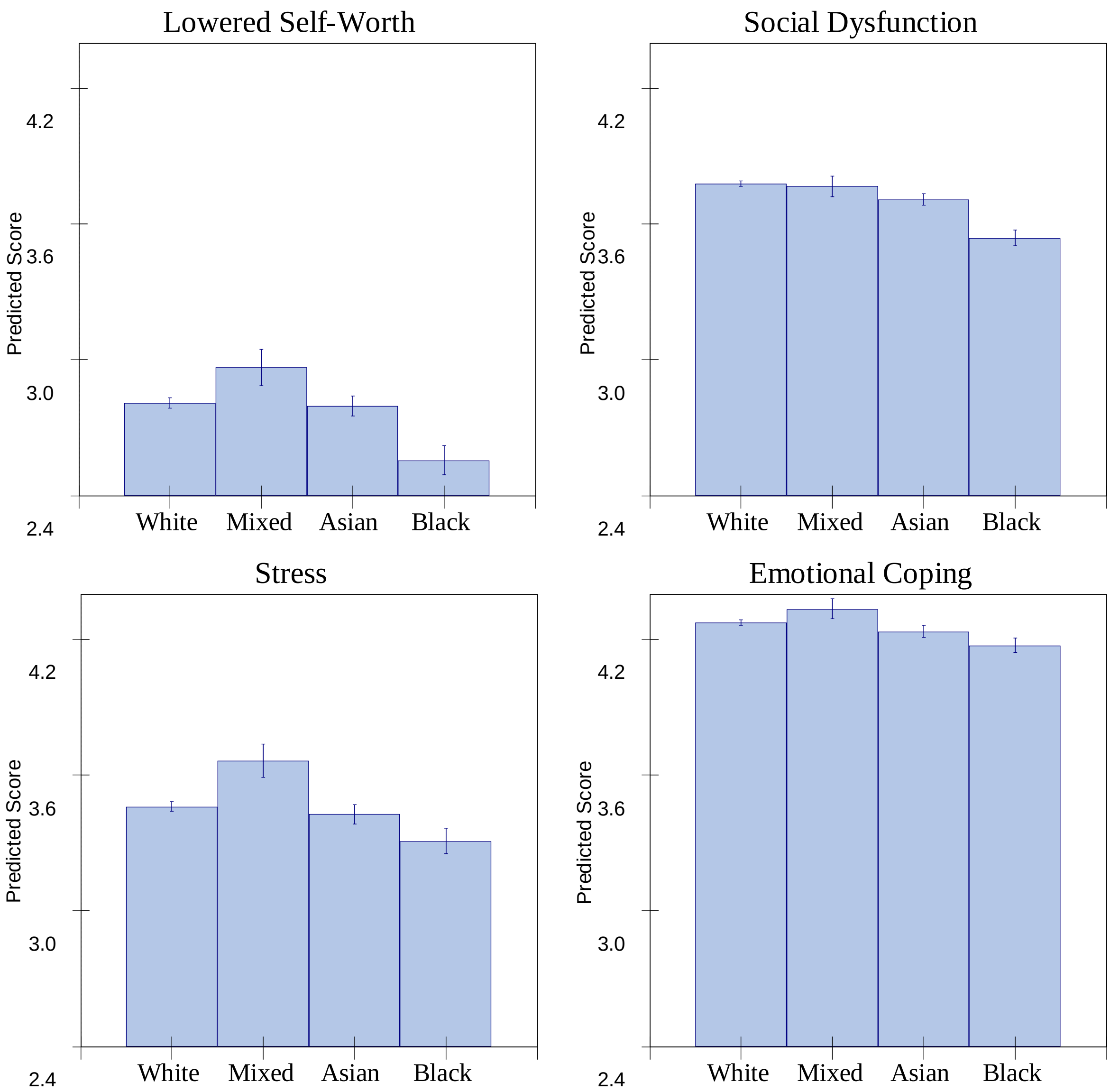


Figure 5.6: Predicted scores across four metrics for different self-reported ethnicity. (White, Mixed Race, Asian, Black)

Figure 5.6 illustrates the differences between self-reported ethnicities for each of the four responses. Individuals who identified themselves as black experience the best mental health as captured by these dimensions. The trends are similar across measures, with the exception of SD. Whites and Asians have marginally worse mental health than individuals identifying as black whilst being statistically indistinguishable from one another for Stress and LSW. Mixed race individuals experience the worst mental health by a statistically significant margin across all metrics except SD. SD bucks this trend with Whites experiencing equally poor mental health to mixed race individuals.

The relationship between ethnicity and mental health is complex, it is often that ethnic minorities experience worse mental health as captured by clinical outcomes, which was initially demonstrated for more easily evidenced psychoses such as schizophrenia (McGovern & Cope, 1987, Harrison et al., 1989). However, this relationship is more difficult to isolate when considering self-rated mental health, although there is evidence that non-white ethnicities have worse mental health (Ross et al. 2000, Jackson-Triche et al. 2000, Hu et al. 2007). The difficulty primarily comes from the problems in isolating a clear effect of ethnicity which is statistically separate from the well-established role of deprivation. It is common to find that whilst ethnicity initially seems to play a large role, this commonly drops when factoring for deprivation due to the increased likelihood of non-white ethnicities to live in deprived areas or experience disproportionate economic disadvantage (Jokela et al. 2013, Oguz et al. 2013, Keyes et al. 2010). Given that here deprivation is controlled for to some extent with education and employment status, results can be expected to be robust to this. These results therefore suggest that there are still differences due to ethnicity, but that these vary depending on which response is considered.

5.4.3.6 Patterning by Employment Category

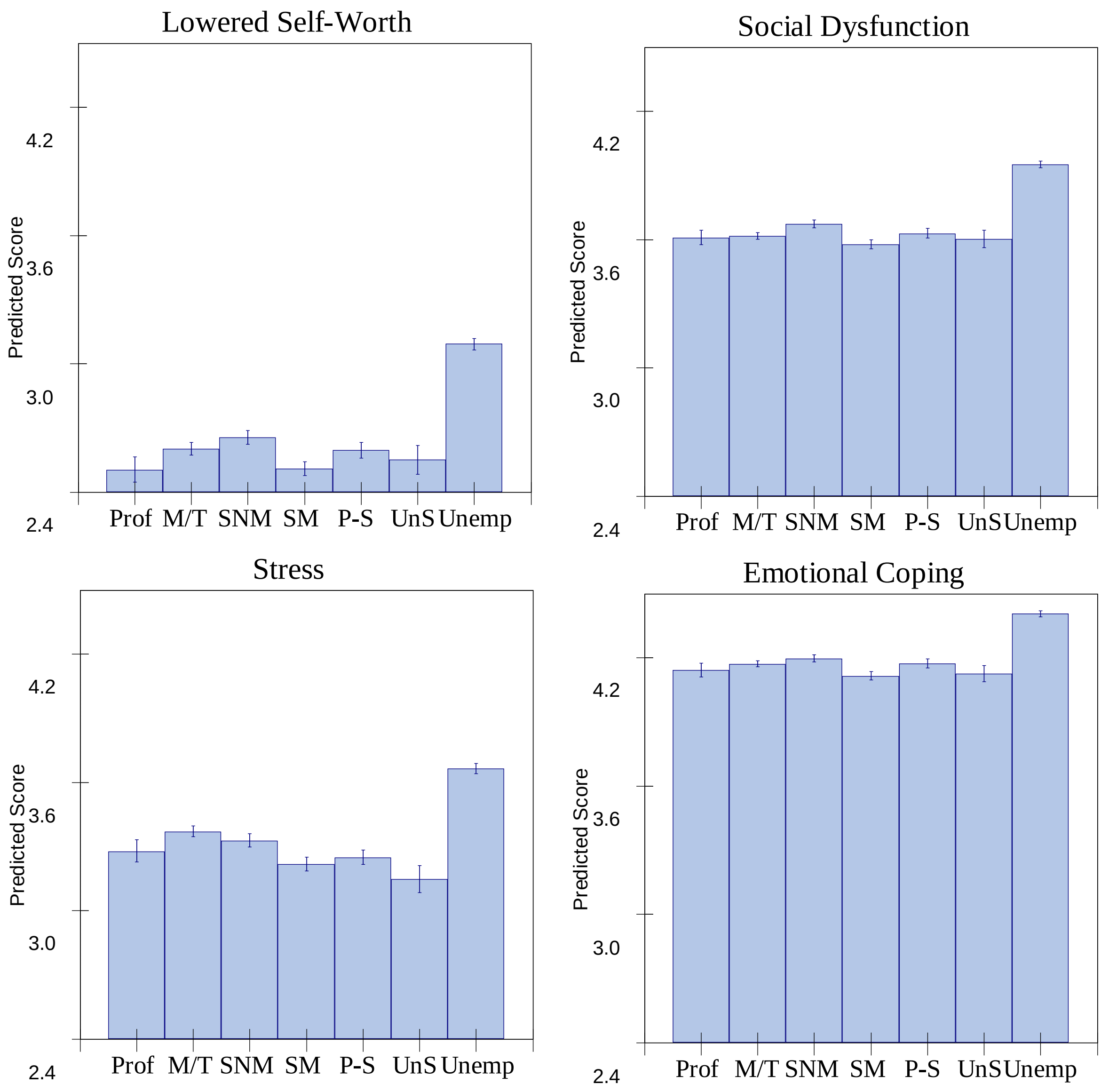


Figure 5.7: Graphs illustrating the effect of different types of Job Classification by Registrar General, on the four dimensions of mental health.
(Professional, Managerial/Technical, Skilled Non-Manual, Skilled Manual, Partly-Skilled, Unskilled, Unemployed).

Figure 5.7 illustrates the differences predicted by Job Classification across the four metrics. There is little differentiation across SD and EC for the various categories of employment and by far the largest difference is that between the employed and the unemployed. This is also the case for the LSW and Stress metrics, however these have more pronounced differences between the various employment types. Unskilled workers have the lowest levels of Stress, where Skilled-Manual and Professional workers enjoy the best self-worth. Across all four metrics, however, the largest gap is between the employed and the unemployed, which corroborates the findings of the majority of the literature. Results show that that the unemployed have significantly worse mental health across all metrics, which is consistent with research into unemployment, has been unemployment to be shown to be predictive of mental distress captured by both positive and negative mental health outcomes (Huppert & Whittington, 2003, van der Noordt et al. 2014, Stewart-Brown et al. 2015).

5.4.3.7 Wave- Year

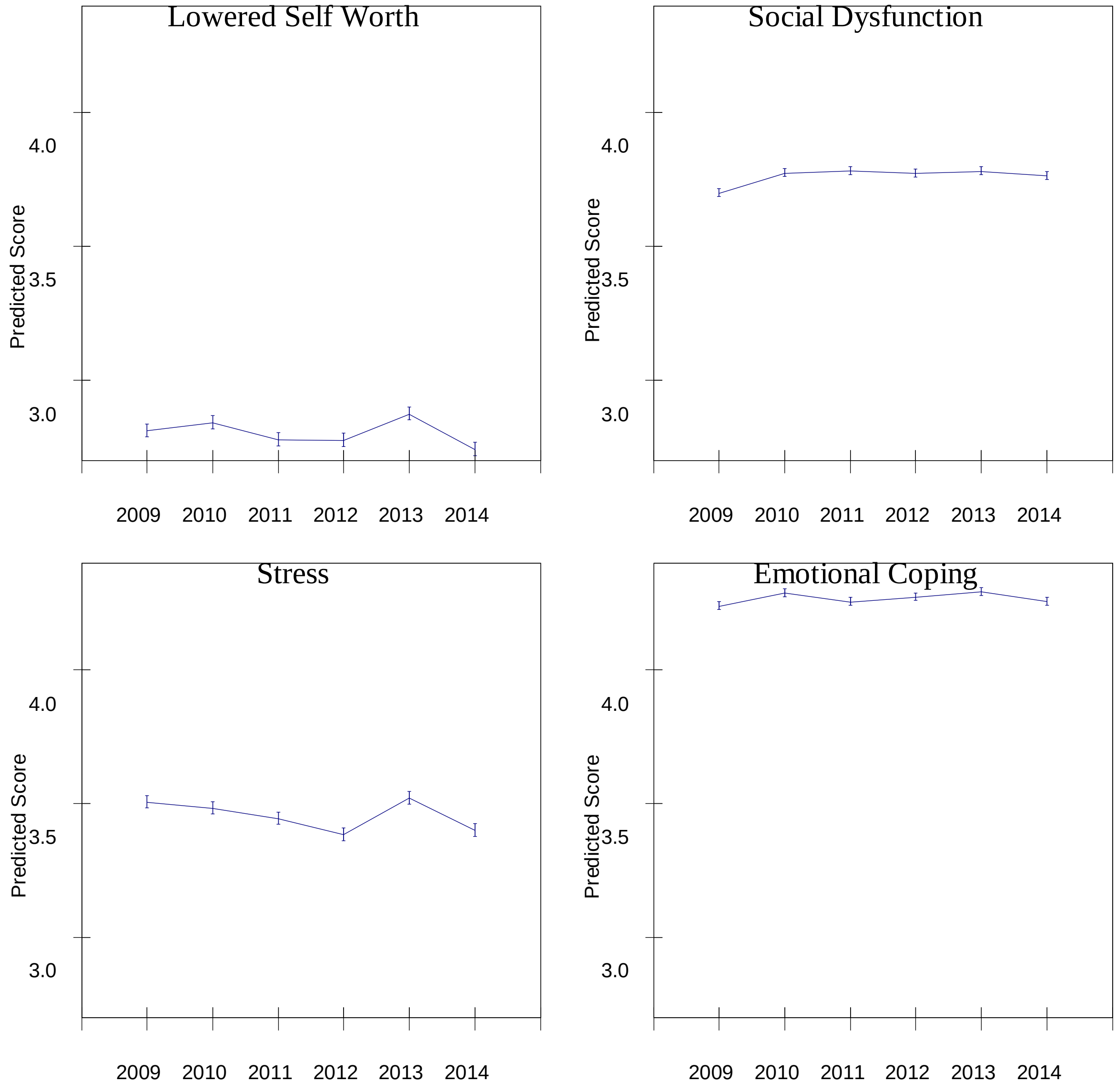


Figure 5.8: Average predicted values for each of the 6 waves across the four responses.
(Each wave constitutes data collected in the two years following the year identified, i.e. Wave 1 constitutes data collected between 2009 and 2011)

Figure 5.8 illustrates the net effect of each year when years are included as a set of categorical predictors for each response. This is useful to isolate if there are any systemic differences year on year that can be taken account of in order to undertake the modelling of different trajectories. This protects against incorrectly ascribing differences to a variable that is truly due to a particularly unhealthy year. There is little change across SD and EC, with aggregate levels remaining very stable, despite a slight increase in SD after 2009. In LSW there is a slight spike in 2010 for Wave 2, preceding a slight decline to a steady state for 2011 and 2012. There is then another larger spike in LSW in 2013 before a decline in 2014. The large spike in 2013 is also seen in the Stress metric, after a period of steady decline in stress since the economic recession. The spike also lasts only one year before returning to pre-spike levels.

5.4.4 Differential Demographic Trends

Having characterised the demographic patterning of the results conditioning on the year of measurement, the final research question asks how differences in mental health experience across each of the dimensions of mental health have differed over the 6 waves. Moreover, it asks not just whether different groups have experienced different net mental health changes over the period, but whether any groups are more volatile or variable in their experience of this. In asking this it is concerned both with differences in trends over the years and differential variability within these aggregate trends. In assessing this, as little structure as possible was imposed on the model, in order to estimate separate terms for each group in each year. This allows the greatest opportunity for differences in demographic fluctuations or trends to be identified. As such, this may introduce fluctuations which are stochastic, and for many groups there may be no significant difference in trends over the study period. To evidence this, confidence intervals are provided around the trend of each group in order to demonstrate the scale and significance of these differences. This was operationalised by specifying interaction terms with each demographic predictor and the categorical wave-term. The polynomial age term was removed and replaced with categorical age-groups to allow separate estimation of coefficients for each age-group in each year.

Results in the following section are grouped not by predictor variable, but by response. This means comparisons of magnitude of effect can be made across predictors for the same response. Graphs are presented in a consistent order for each response, reading left to right, top to bottom, Sex, Age, Marital Status, Ethnicity, Educational Attainment and Job classification. The categories and labels for these graphs are as in Table 5.4[34].

5.4.4.1 LOWERED SELF-WORTH

Figure 5.9 shows the differential trends in LSW in the years following 2009 for 6 demographic predictors. Across all predictors there is evidence of the peaks in mental distress in years 2010 and 2013 that was visible in Figure 5.9, however these peaks seem to be markedly stronger for some groups. There is again evidence of the dissimilarity between males and females evidenced in Figure 5.2. Immediately following 2009 there is evidence of a widening of the gap between the mental health of females and males. There is evidence of this divergence in the years between 2010 and 2012, with males steadily improving and females staying static and slightly worsening, but males were more strongly affected by the peak in 2013, bringing the difference back to comparable levels as in 2010. There seems to have been a marked improvement across both sexes in the 2014 wave, improving self-worth for both sexes.

There are much more noticeable differences between groups based on age. There is evidence of the same relationship illustrated in Figure 5.3, with the middle-aged exhibiting worse mental health across all years considered. Within this broad relationship however, there is a marked worsening in self-worth for the youngest group, those aged under 25 in 2009. Whilst the rest of the age groups either improve or stay static in 2011, there is a considerable worsening in self-worth for under 25s, with a slight respite in 2012 before worsening again in 2013. The under 25s illustrate the same improvement in 2014 that is present across groups, but the differences between age groups in 2014 are markedly dissimilar from those in 2009. The over

[34] As in the previous section, Lowered Self-Worth, Social Dysfunction and Emotional Coping are abbreviated to LSW, SD and EC respectively.

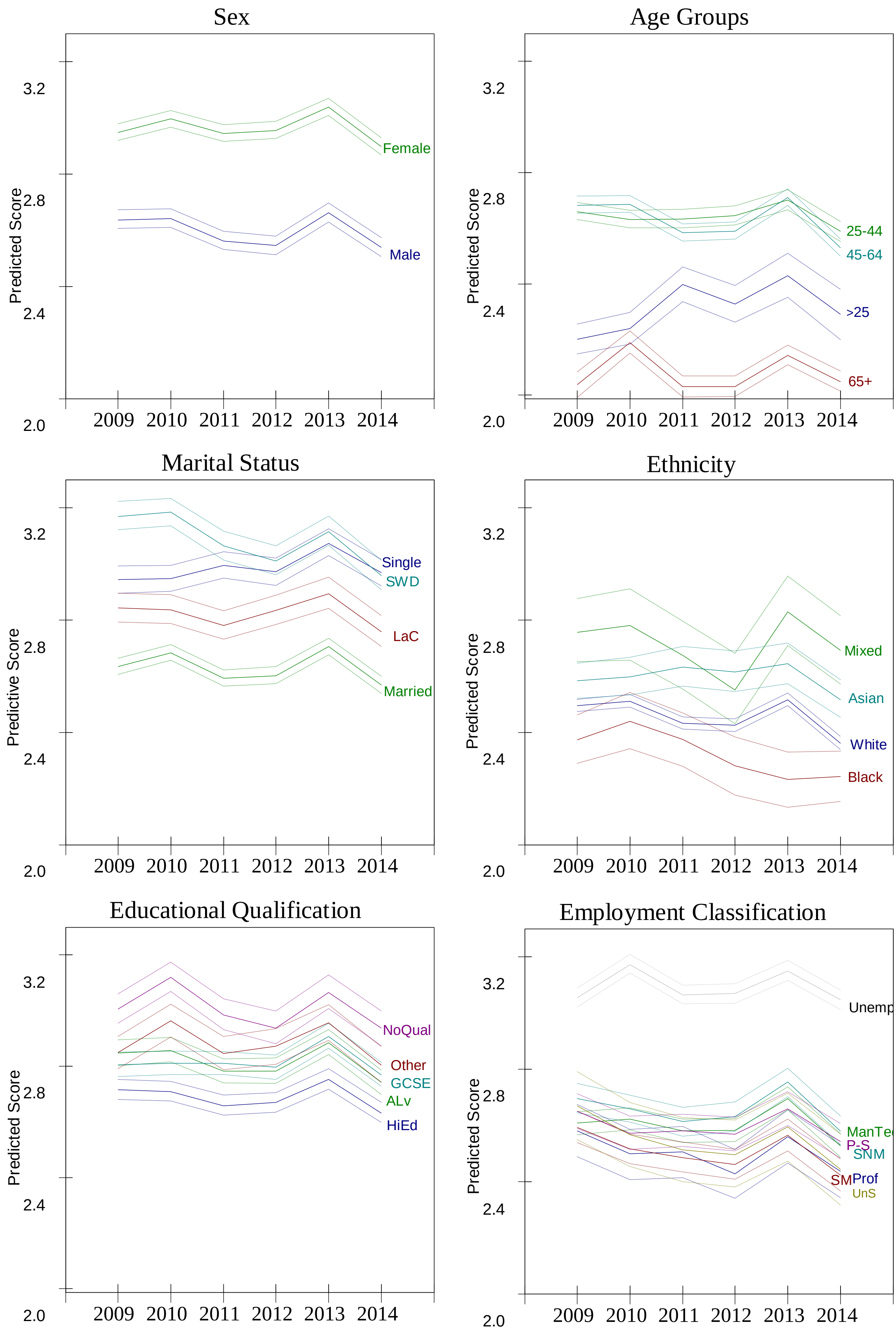

Sex
Predicted Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
Female
Male
Age Groups
Predicted Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
25-44
45-64
>25
65+
Marital Status
Predictive Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
Single
SWD
LaC
Married
Ethnicity
Predicted Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
Mixed
Asian
White
Black
Educational Qualification
Predicted Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
NoQual
Other
GCSE
ALv
HiEd
Employment Classification
Predicted Score
3.2
2.8
2.4
2.0
2009 2010 2011 2012 2013 2014
Unemp
ManTe
P-S
SNM
SMProf
UnS

Figure 5.9: Graphs illustrating predicted Lowered Self-Worth values by year across demographic predictors. Left to right, top to bottom these are Sex (Male, Female), Age Band (<25, 25-44, 45-64, 65+), Marital Status (Single, Married, Separated/Widowed/Divorced, Living as Couple), Ethnicity (White, Black, Mixed, Asian), Highest Educational Attainment (Higher Education, A-Level, GCSE, Other Qualification, No Qualification) and Job Classification (Professional, Managerial/Technical, Skilled Non-Manual, Skilled Manual, Partly-Skilled, Unskilled and Unemployed).

65s experienced the largest spike in lowered self-worth in 2010, but then returned to pre-2010 levels within a year.

The same broad trends are illustrated for Marital Status as illustrated in Figure 5.4. Married individuals exhibit the best mental health, with individuals living as couples experiencing better mental health than the single, separated, widowed or divorced. There is a noticeable convergence between the SWD and the singles in 2011. This was, again, due to singles disproportionately worsening in 2011 relative to the improvements in mental health for the other groups. The mechanism for this is likely similar to that of the youngest group across the waves younger individuals are disproportionately single. From 2011 the singles and SWDs are statistically indistinguishable. There is evidence of the spike in 2013 across all groups, most notably in the married individuals who have otherwise remained relatively static over the period.

Most individuals in the study are white, as illustrated by the tight confidence interval. Despite the low number of participants across other ethnicities there are still significant differences between groups. Asians experience significantly worse self-worth than whites in years following 2010, although notably the spike that is seen across nearly every other group in 2013 is not present for Asians or Black individuals. The trajectory for mixed race individuals is very volatile, somewhat due to the low sample size, but illustrates a marked improvement in the years running up to 2013 before the largest negative spike of any group in the 2013 wave. Black individuals consistently show the best mental health, although are statistically indistinguishable from whites in the years 2010 and 2011.

Individuals with university level education exhibit significantly better self-worth over the entire time period, and (along with GCSE achievers) do not see the same spike in LSW in 2011 that other groups experience. There is no significant difference in any of the years between the self-worth of those achieving GCSEs or A-Levels. Those with no qualifications experience the worst mental health, although it is statistically indistinguishable from those with other qualifications for all years.

The largest difference across any of the predictors, as expected from Figure 5.7 is between that of the unemployed and the employed, with the unemployed experiencing considerably worse self-worth across the whole period. There is little significant difference between the various categories of employment beyond the significant differences between the Skilled Manual and Professional workers who experience the best self-worth, and the Skilled Non-Manual who seem to experience the worst self-worth of the employed. Interestingly the worsened self-worth that seems to have been exhibited in 2011 across most other groups is not evidenced strongly amongst the employed, however it is strongly evidenced among the unemployed.

5.4.4.2 SOCIAL DYSFUNCTION

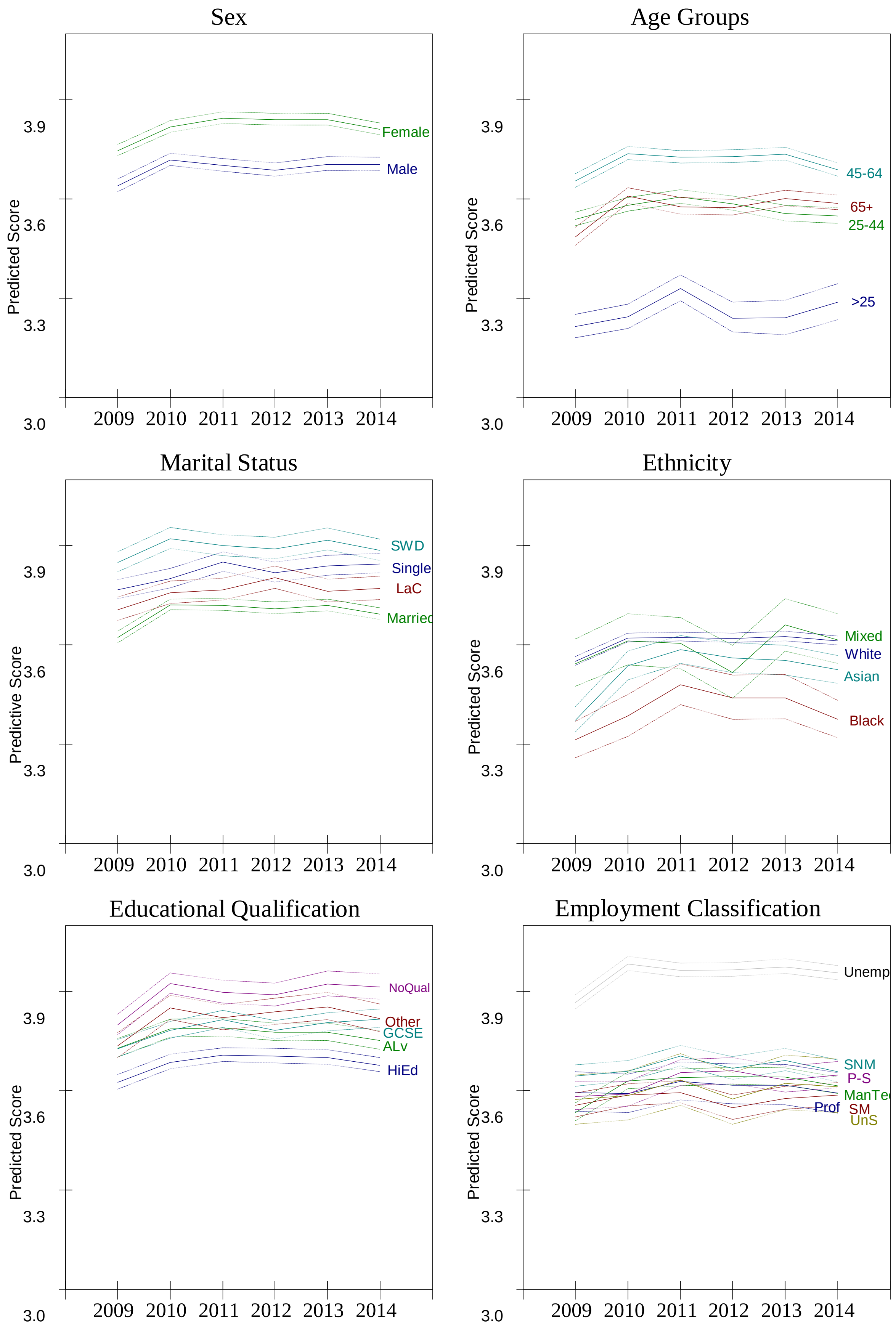


Figure 5.10: Graphs illustrating predicted Social Dysfunction values by year across demographic predictors. Left to right, top to bottom these are Sex (Male, Female), Age Band (<25, 25-44, 45-64, 65+), Marital Status (Single, Married, Separated/Widowed/Divorced, Living as Couple), Ethnicity (White, Black, Mixed, Asian), Highest Educational Attainment (Higher Education, A-Level, GCSE, Other Qualification, No Qualification) and Job Classification (Professional, Managerial/Technical, Skilled Non-Manual, Skilled Manual, Partly-Skilled, Unskilled and Unemployed).

The graphs in Figure 5.10 illustrate the changes in SD for different demographic groups. As Figure 5.8 exemplified, there is little aggregate change over the time period, however there is clear differentiation in the experience of mental health between groups. There is evidence as in Figure 5.2 of the difference between male and female mental health. The difference between the two sexes is less here than illustrated for LSW in Figure 5.9, but this is to be expected given the results in Figure 5.2. Despite this, there are clear differences in trends. SD worsened for both sexes in 2010, however improved in the following year for males, and worsened for females. This divergence is more marked than in Figure 5.9 for LSW. There is also evidence of convergence in 2014, where male social dysfunction worsened, and female social dysfunction improved.

There are more static age trajectories for social dysfunction than for LSW. Given this there are much more defined peaks in mental distress for the youngest and eldest categories. The over 65s experience a sharp decline in social function in 2010, and the under 25s experience a similarly large decline in social function in 2011. Both these peaks are relatively rapidly overcome, and levels return to their pre-peak norms. This fits with the individual level VPCs given in Table 5.7, which suggest that social dysfunction is more transient than LSW and Stress.

Married individuals still enjoy the best social functioning, although the difference between the married and those living as couples is far less marked in the years before 2011 for SD than LSW. Post 2011, those living as a couple experience a spike in mental distress that pushes them into the same level of distress as singles. The separated, widowed and divorced group does not have the same level of improvement relative to the singles in 2011, and they experience worse mental health by a significant margin for all years except 2011 and 2014.

White individuals see almost no change over the period in SD. The other ethnic groups are far more volatile, with the most volatile being those of mixed race. Individuals identifying as black experience strongly worsening social functioning from 2009 until 2011 before improving steadily until 2014. Asians experience a large spike in social dysfunction in 2010 before levelling off, and mixed-race individuals again illustrate a considerable improvement in 2012 before worsening strongly in 2013.

The most highly educated again illustrate the best mental health as captured by social dysfunction. There is evidence of a slight worsening among this group until 2011 before levelling off and staying static until 2014. Those with no qualifications see the worst social functioning in all years bar 2012 and 2013, and do not see the same degree of improvement in social functioning that is common across all other groups in 2014. Again, those who achieve A-Levels and those who achieve GCSEs are not easily distinguished beyond having worse mental health than those who attended university.

There is again little distinguishing between the trends of the various employment categories. Managerial/technical workers experienced a disproportionate worsening in social functioning in 2010 relative to other groups and did not recover. The largest difference, again, is between the unemployed and the employed, with the unemployed experiencing by far the worst social functioning. They also had a far greater worsening of their social functioning in 2010, and this never recovered, staying static at this elevated level through until 2014.

5.4.4.3 STRESS

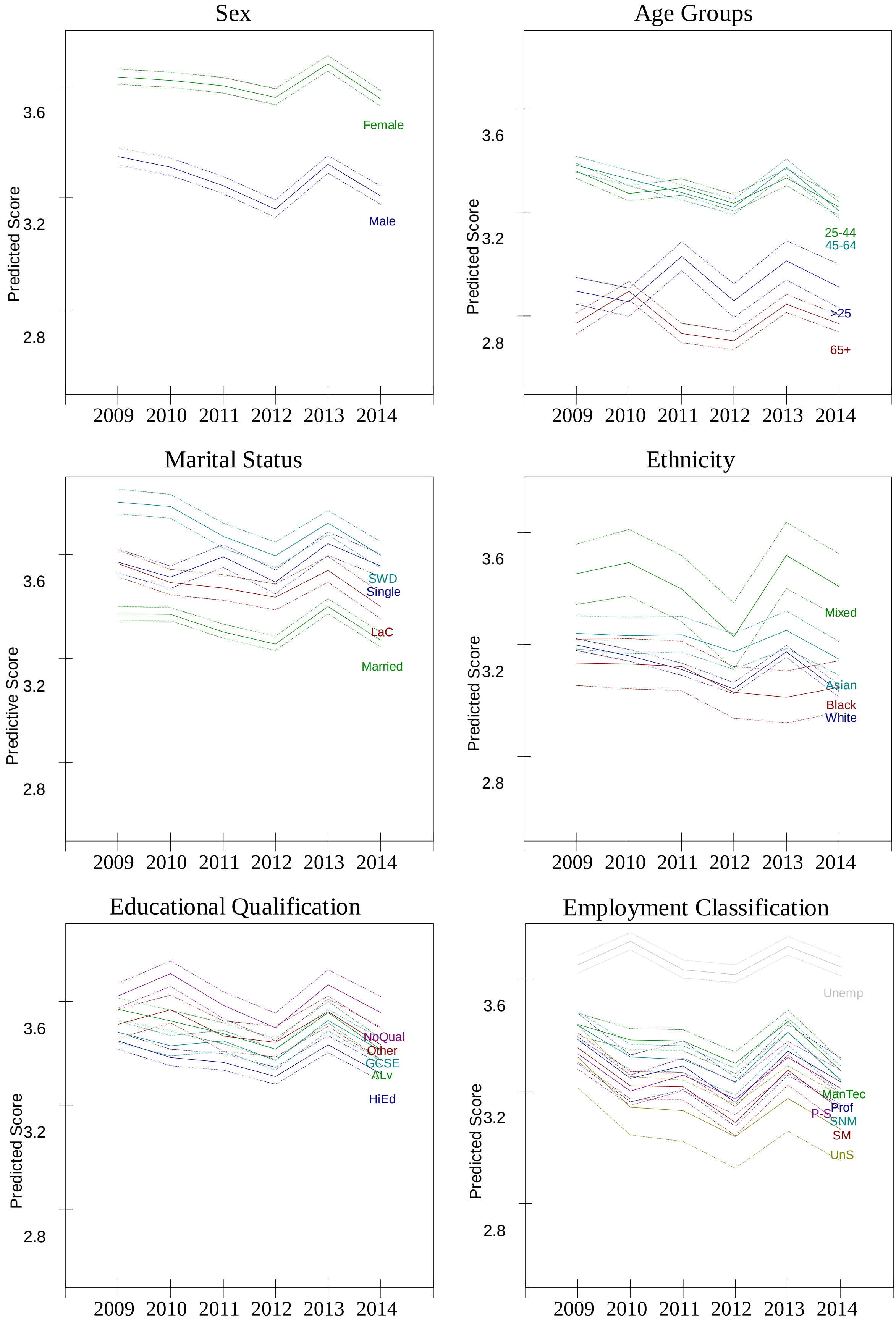

Sex
Predicted Score
3.6
3.2
2.8
Female
Male
2009 2010 2011 2012 2013 2014
Age Groups
Predicted Score
3.6
3.2
2.8
25-44
45-64
>25
65+
2009 2010 2011 2012 2013 2014
Marital Status
Predictive Score
3.6
3.2
2.8
SWD
Single
LaC
Married
2009 2010 2011 2012 2013 2014
Ethnicity
Predicted Score
3.6
3.2
2.8
Mixed
Asian
Black
White
2009 2010 2011 2012 2013 2014
Educational Qualification
Predicted Score
3.6
3.2
2.8
NoQual
Other
GCSE
ALv
HiEd
2009 2010 2011 2012 2013 2014
Employment Classification
Predicted Score
3.6
3.2
2.8
Unemp
ManTec
Prof
P-S
SNM
SM
UnS
2009 2010 2011 2012 2013 2014

Figure 5.11: Graphs illustrating predicted Stress values by year across demographic predictors. Left to right, top to bottom these are Sex (Male, Female), Age Band (<25, 25-44, 45-64, 65+), Marital Status (Single, Married, Separated/Widowed/Divorced, Living as Couple), Ethnicity (White, Black, Mixed, Asian), Highest Educational Attainment (Higher Education, A-Level, GCSE, Other Qualification, No Qualification) and Job Classification (Professional, Managerial/Technical, Skilled Non-Manual, Skilled Manual, Partly-Skilled, Unskilled and Unemployed).

Figure 5.11 illustrates the differential trends in Stress for different demographic groups. As in Figure 5.8 there is a clear trend across the years of declining stress until 2012 followed by a large spike in stress in 2013 before a recovery in 2014. This trend seems to be more marked for males than females, although this is largely because females did not experience the reduced stress levels in the years preceding 2013 that males enjoyed. The difference between sexes is largest in 2012, where the spike in stress in 2013 more strongly affected men and thus moved the two closer together.

The spike that has been noted in 2011 for the youngest age group in LSW and SD is even more pronounced in stress. Between 2009 and 2010 all age categories improved apart from over 65s who experienced a peak in stress in 2010. In 2011 however, the under 25s experienced an even larger spike in stress levels. It is not the aim here to speculate on the cause for this, but this does coincide with the increase in tuition fees that was proposed in 2011, which would have disproportionately affected this age group. This is immediately rectified in 2012 with all groups experiencing a reduction in stress before a spike across ages in 2013. This spike seems to be larger amongst the 45-64 year-olds than other age bands. All ages saw an improvement in stress in 2014.

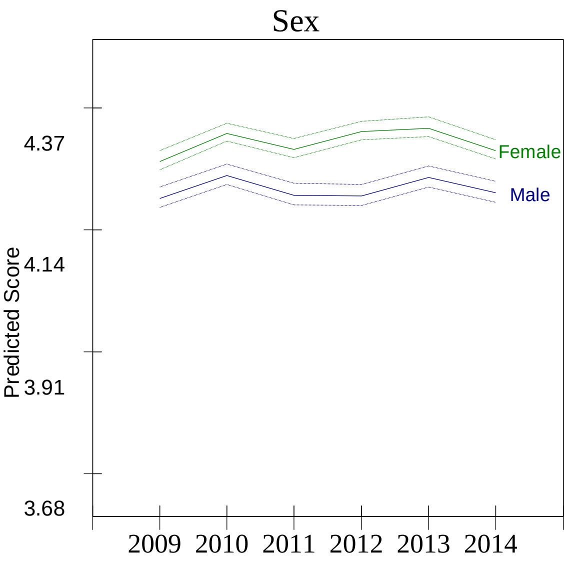
Trends in stress by marital status look fairly similar to those in LSW, however there are some differences. Married individuals enjoy the best stress levels by a significant margin across all 6 waves. There is no difference in stress between individuals who are single and individuals living in a couple except in 2011 when singles experience a large spike in stress, and 2014 where singles do not experience the improvement in mental health that other marital status groups do. Similar to LSW, the separated/widowed or divorced individuals experience significantly worse stress levels before 2010, before becoming indistinguishable from singles.

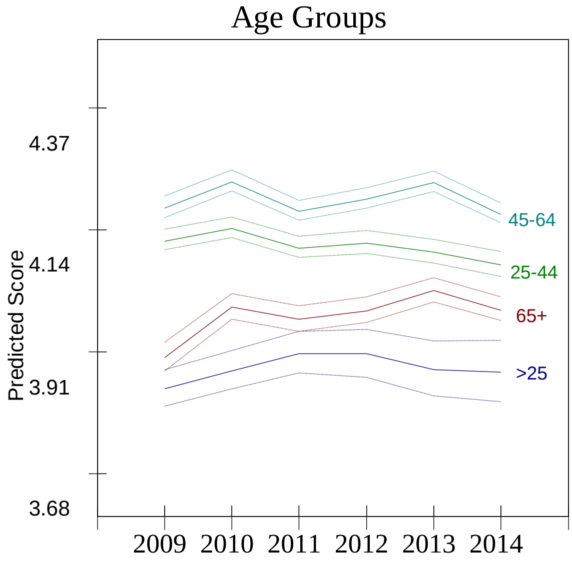
Except for mixed race individuals, non-whites do not tend to experience significantly different levels of stress to whites. Mixed race individuals experience significantly worse stress levels than whites at every wave, and experience the worst overall with the exception of Asians in 2011 and 2012, where they do not differ significantly. Blacks and whites experience almost identical levels of stress until 2012, before whites suffer the same spike in stress that almost all groups experience, except for those identifying as black.

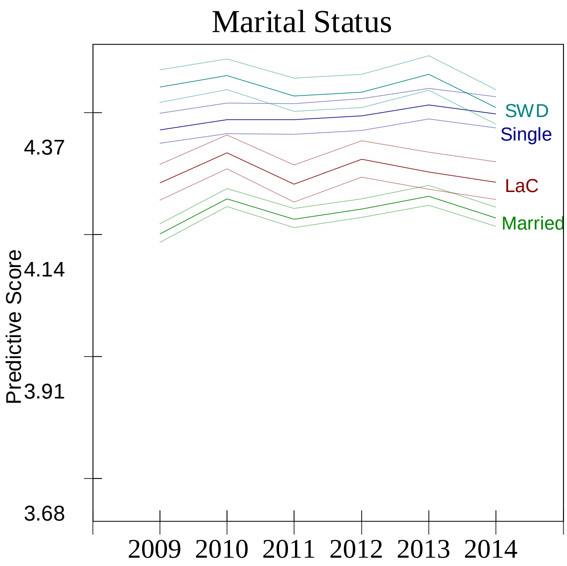
There are few significant differences in the trends in stress for individuals based on Educational attainment. The most noticeable difference is that the spike in stress experienced in 2011 seems to, again, only affect those with no qualification or other qualifications. There is some evidence of a dose-response relationship with education in the later years, but until 2012 there is no significant difference in stress levels between those who achieved A-Level and those who went on to university.

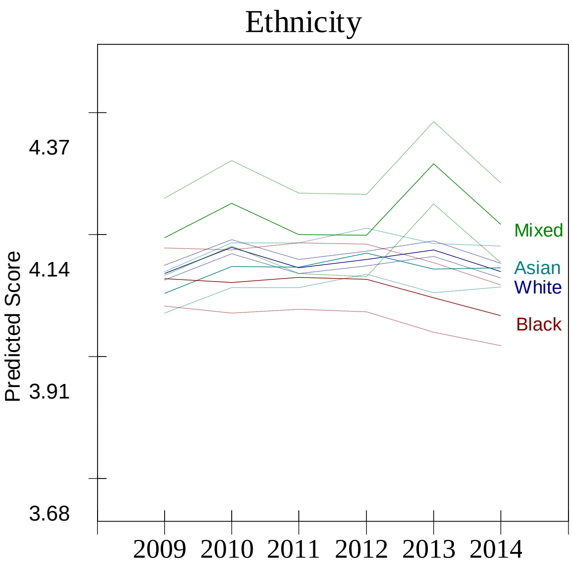
The largest difference in stress levels is, perhaps unsurprisingly, between the unemployed and the employed. The unemployed have consistently elevated stress levels across all years, and do not seem to enjoy the same degree of alleviation of this stress that other employment groups experienced in the years running up to 2012. Perhaps because of this lack of improvement in 2012, the jump in stress experienced across other categories is less pronounced in the unemployed. There are not particularly large differences in the temporal trends of the different categories of employment beyond the binary of being employed or not.

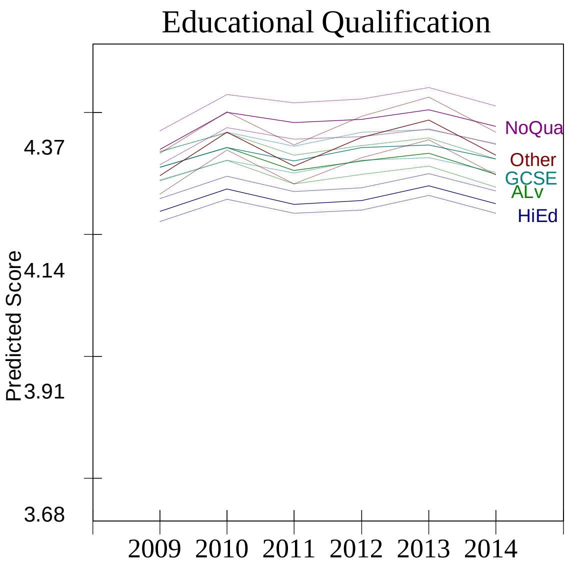
5.4.4.4 EMOTIONAL COPING











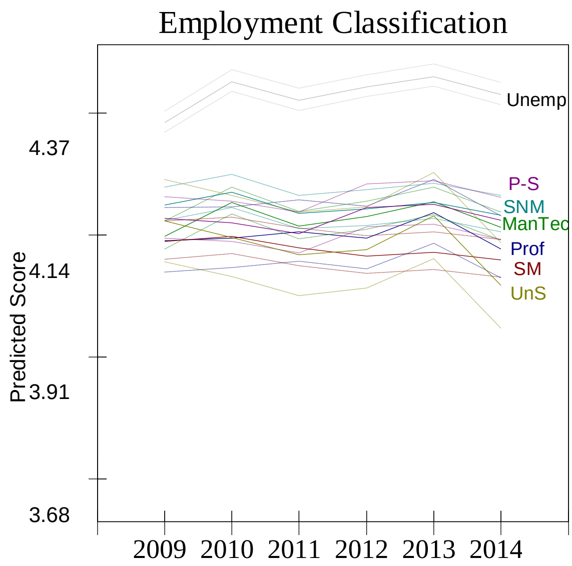


Figure 5.12 Graphs illustrating predicted Emotional Coping values by year across demographic predictors. Left to right, top to bottom these are Sex (Male, Female), Age Band (<25, 25-44, 45-64, 65+), Marital Status (Single, Married, Separated/Widowed/Divorced, Living as Couple), Ethnicity (White, Black, Mixed, Asian), Highest Educational Attainment (Higher Education, A-Level, GCSE, Other Qualification, No Qualification) and Job Classification (Professional, Managerial/Technical, Skilled Non-Manual, Skilled Manual, Partly-Skilled, Unskilled and Unemployed).

Figure 5.12 gives the predicted values for EC across the demographic predictors. Differences between sexes are far less marked in EC than in other responses, demonstrating the temporal consistency of the difference shown in Figure 5.2. It is important again to acknowledge that whilst the other scales have a more clearly defined “negative” end of the scale, EC refers to the negative emotion and psychological distress of individuals but is also positively associated with feeling capable of decision-making. The trend broadly follows the twin-peak trend that is seen in Figure 5.8, with spikes in 2010 and 2013. There is little overall change over the time period, however there are fluctuations. There is a spike consistent across both sexes in EC in 2010, and a slight divergence in 2012 where females scores increase disproportionately to males, but other than this the trends are very similar.

The patterning of EC by age groups over time shows more consistency in temporal trends than the other responses. Over 65s have a clear and fairly steady increase in their EC over the time period, with slight peaks in 2010 and 2013. The youngest individuals, the under 25s, show an increase over the first three years before levelling out and staying static apart from a slight decline between 2012 and 2013. The 25-44 age group seems to be declining in its tendency towards being stoic over the years since 2010, whilst the 45-64 group fluctuates with peaks in 2010 and 2013 but does not show an aggregate difference over the time period, staying consistent in their level of EC. The increases in the over 65s and under 25s suggest that these groups are feeling more negative but perhaps becoming more positive about their decision-making capacity in the face of this emotional distress, although looking at the results for the other responses this is not a change that is easy to characterise more broadly.

Single people are the most consistent of any group in their levels of EC with only a very slight increase over the six waves. Married individuals have the least distress and also perhaps least trouble correlating that emotional distress with reduced decision-making capacity, and experience significantly lower scores than all other marital classifications across all years, with the exception of a slight overlap of confidence intervals in 2013 with those living as a couple. Those living as a couple are the most volatile, with peaks in 2010 and 2012, but both rapidly returning to previous levels. Again, this is consistent with the idea that EC is similar to SD in that it is determined more strongly by occasion than individuals.

There is almost no discernible difference between the trends of individuals based on their ethnicity. The commonality of the EC construct seems to suggest that this metric behaves differently to broader mental health trends which are commonly discussed in the literature as having large differences based on ethnicity.

There are also very few significant differences in EC between individuals with different levels of educational attainment. Those who attended university score lowest by a significant margin across all years. The trends between the remaining categories is very similar with the exception of those with other qualifications who experienced a larger peak than the others in 2010, followed by a unique dip in EC in 2011 before a steady rise until 2013 and a decline in 2014.

The unemployed are again the most notable category for Job Classification differences. They are increasingly likely to experience distress, and perhaps deem their decision-making capacity as unaffected by negative emotion over the period, with peaks in 2010 and 2013. The employed categories do not seem to experience significantly differing trends across the time period, although it is notable that professionals are in the middle of the pack of the employed categories, where they have tended to be on the lower end of the other scores.

5.5 Conclusions

This Chapter presents potentially the most comprehensive model of changing UK mental health as captured by the GHQ-12 to date. Stratifying mental health into four separate but related categories each constructed from the GHQ-12, longitudinal models are specified to analyse demographic and temporal patterns in mental health in the years following 2009. These models are constructed in an exploratory manner to maximise the predictive capacity of the model in explaining variation across the four modelled responses. Moreover, the final model structure allows the partitioning of variance into different structural levels, comprising area-level, individual-level and occasion-level variation. This allows inquiry into the degree of temporal consistency across these responses. Furthermore, due to the specification of demographic predictors in the model, it is possible to analyse the net effect of these demographic predictor variables over time, taking account of year-on-year variation as well as the geographical dependency introduced by clustered sampling. Finally, the inclusion of longitudinal interaction terms allows the estimation of separate trends in *each* of the mental health constructs for each of the demographic groups.

5.5.1 Temporal Consistency of Mental Health

The variance/covariance structure of the model gives the degree to which the responses are correlated at each level. Correlations were highest at the PSU-Level, suggesting geographical consistency between mental health dimensions whereby for example stressed areas are likely to be areas of Low Self-Worth. There were some interesting differences between occasion level and individual level correlations for Social Dysfunction. This demonstrates that individuals struggling with high levels of Social Dysfunction were more likely to be stoic. However, if there was an *occasion* where individuals are finding social functioning challenging than usual, this is more likely to be associated with a period of high levels of stress.

Partitioning of variance illustrated that Lowered self-worth was mostly determined by between-individual differences, but all three other responses were most variable at the within-individual between-occasion level. VPCs at PSU-level were modest, at 2% for Lowered Self-Worth and Stress, and ~1% for Social Dysfunction and Emotional Coping, however these are not negligible given that they are net of individual and temporal variation. This demonstrates the worth of area level investment in mental health improvement. Carrying this work forward, investigation of negative area-level residuals would highlight which areas would benefit most from structural investment into mental health improvement. Variance partitioning indicated that the specific occasion was twice as important as the individual for predicting Social Dysfunction and Emotional Coping, where for Stress and Self-Worth occasion is equally as important as the individual. This suggests that Self-Worth and Stress are more stable within an individual than that individual's evaluation of their performance in a wider social setting, and their tendency to be stoic. Across all measures bar Lowered Self-Worth, within-individual between-occasion variation was the most important structural level. This reads as a heartening message about the non-permanence of negative mental health. The most important factor in an individual's mental health is *not* the individual themselves – but the specific occasion. This suggests that notions of "mentally unhealthy individuals" should potentially be reframed as "mentally unhealthy circumstances", particularly when considering aspects of social functioning, and perceived competence in light of negative mental experience This is particularly pertinent given the acknowledged negative impact of stigma associated with the individualisation of mental distress. In reality, there is clear capacity to change through time, and time of measurement is more important than the individual. This has large-scale implications for policy, which should target investing in resources known to buffer mental health across individuals, such as education, mental health awareness and measures to combat isolation, loneliness and discrimination.

5.5.2 Demographic Patterning

Overall the results for the net effect of demographic predictors were similar to those observed in Chapter 4. Within this there were noticeable dissimilarities between responses, with constructs differing markedly in their variability with respect to demographic predictors. Differences were greater across the board between categories when considering the Lowered Self-Worth and Stress responses, with Social Dysfunction and Emotional Coping being more consistent across groups. Overall females, middle-aged individuals, the separated/widowed/divorced, those with no qualifications, mixed race individuals and the unemployed tended to display the highest scores across the four responses. The largest differences are between the ill and well are between the unemployed and professional or skilled manual workers, and those in their mid-40s and those in their late teens respectively.

There are interesting dissimilarities across responses for males and the elderly across metrics. In Lowered Self Worth and Stress, these groups showed better mental health. However, this difference was far smaller on the Social Dysfunction and Emotional Coping metrics. This has been suggested to potentially be a product of their tendency to under-report emotional suffering when faced with negatively phrased questions. Separating marriage and cohabiting showed that living as a couple is almost as protective for mental health as being married, but married individuals still enjoyed the best mental health. Differences based on education were small but significant, with those with higher education enjoying the lowest scores across metrics. Ethnicity proved to be complex to interpret, with blacks enjoying the lowest scores across metrics, and mixed-race individuals giving the highest scores. Whites and Asians experienced comparable mental health across all metrics bar Social Dysfunction, in which whites demonstrated higher levels of distress, comparable to that of the mixed-race individuals. The largest difference predicted between two groups across any of the responses was between the unemployed and the employed, with the unemployed experiencing by far the worst mental health across all metrics, with this difference most notably evidenced in their levels of Self Worth and Stress.

5.5.3 Demographic Trends

There are numerous small differences in mental health trajectories for different demographic groups across the period. The net effect across demographics was of 2010 and 2013 being particularly poor years for mental health, especially so for Stress. The most striking difference is between under-25s and all other groups in the wave commencing 2011, where across all dimensions under-25s saw a large spike in mental distress. This could speculatively be linked to the rise in tuition fees in this period, which would disproportionately affect the youngest members of the panel. This spike was echoed in singles, who experienced a disproportionate increase in mental distress in the same year, although this is likely to be due at least in part to their correlation with one another. Similarly, there was a large spike in mental distress amongst the over-65s in 2010. Overall there was a notably low-stress year in 2012, however this was most strongly experienced by males, whites and married individuals. Mixed Race individuals also seem to be far more volatile over time across all metrics than other ethnicities, especially whites. This is most clearly evidenced in the Stress response, with a larger effect of the low stress year in 2012 followed by a far higher peak in 2013.

Overall this final chapter highlights the very real differences in mental health experienced by different demographic groups. It also clearly demonstrates the benefit of advocating nuance when interpreting traditional mental health metrics. Whilst correlated, especially at aggregate levels, it is clear that there are very different processes operating for each of these responses at the individual- and occasion-levels. The most striking finding is the difference in temporal variability of these constructs within individuals, with high levels of Lowered Self-Worth and Stress being more consistent over time, and Social Dysfunction and Emotional Coping being much more variable within individuals. These markedly different constructs illustrate markedly different patterns and make up part of a much broader argument for increased sensitivity and nuance in interpreting mental health screening questionnaires. Whilst there is use in aggregate interpretation for understanding broad scale changes over time, there is clearly more going on than initially meets the eye, and this should be addressed in the methodological choices of researchers interested in geographical, temporal and demographic differences in mental health.

5.5.4 Limitations

Whilst this is a powerful and far-reaching chapter in terms of investigative purpose, the findings here should be considered in light of several key limitations. These are grouped into; response limitations, panel limitations, and modelling limitations, and are detailed below.

Firstly and perhaps most importantly, is the issue with self-rated mental health measures. With all investigations into self-rated mental health it is difficult to distinguish what constitutes a genuine difference in an underlying mental state, and what constitutes a difference in propensity to respond to a certain type of question. For instance, stoicism is statistically identical to experiencing less distress. This is an issue that it is not possible to directly address methodologically beyond increased awareness of mental health at a societal level. If individuals are more aware of mental illness and mental health as concepts, and there are efforts to reduce stigma in mental health discussion, this will not only improve their mental health on an individual level but may also improve it at a societal level as measures of mental health will become more accurate. Notions such as stoicism and models of relative social comparison, which have been explored in this thesis are fundamentally confounding in self-reported mental health research, and incredibly difficult to isolate. There are considerable differences in the findings of investigations into self-reported mental health, and the findings of investigations into clinical outcomes such as suicide, as referenced in Chapter 1. Given these differences it seems there is still much work to be done here in making mental health something that all groups feel equally comfortable discussing.

5.5.4.1 Response limitations

Across the constructs modelled here it is difficult to directly contrast observed difference in responses. Whilst the absolute range of the calculated constructs is identical from 0-10, the distribution within this range is quite different, as illustrated in Figure 5.1. This was the justification for the grouping of trends by response in section 5.4.4, as comparison on the same scale is impractical across responses. Caution should be exercised when trying to directly compare changes across groups, across responses, for this reason. Similarly, the numerical differences predicted by different demographic variables are likely to mean different things for each response. This is due to them being composed of separate items in the GHQ-12, and in the case of Emotional Coping being composed of fewer items. This would lead us to expect the distribution of the 0-10 standardised Emotional Coping construct to be slightly broader than the other constructs, however in Figure 5.1 we see this is not the case.

In specifying responses which are generalised over the entire time-period this chapter assumes that the constructs identified in Chapter 2 and modelled here are temporally stable. That is, the factor structure underpinning the GHQ-12 does not change over the 6 waves. Whilst other proposed factors underpinning the GHQ-12 have been demonstrated to be relatively consistent through time, this does not mean that the structure identified here will necessarily be. If the meaning of the questions changed for respondents over the time-period, then this will likely bias results. Further study should aim to carry out ESEM on GHQ-12 data gathered at a later wave in order to test the temporal consistency of the constructs identified in wave 1.

5.5.4.2 Panel Limitations

As with all longitudinal and cohort studies, Understanding Society suffers from ongoing cohort attrition as the panel develops. Although this has been suggested to be less of an issue for Understanding Society than in similar panel and cohort studies, this still carries the capacity to bias results (Knies, 2017). Whilst empirical studies must assume that missing data is at random, for inference to a general population, in longitudinal studies this is often not the case. Table 5.2 demonstrated the demographic patterning of attrition, displaying the bias towards whites, the highly educated, highly employed and female respondents. Whilst Table 5.2 actually overestimates the effect of attrition, as it only includes individuals who remained in for all waves rather than those who exited and re-entered the panel, it is important to acknowledge the possibility that conclusions here are biased to some degree by this non-random attrition.

As alluded to in the introduction to this chapter, longitudinal analysis attempting to explicitly characterise age and cohort effects is often problematic. This is due to the exact confounding between age, period and cohort (APC). Subtracting cohort from period necessarily produces age (Bell and Jones, 2013). If at any stage a researcher knows two of these, they inevitably will know the third. Therefore, it is logically impossible to empirically discern the effects of one of these entirely independently of the others (Bell and Jones, 2018). This means that the trends ascribed to age-group cohorts in the results here, could equally be a product of the year of measurement, or of the age of individuals in any given year. Whilst this is problematic if trying to ascertain the underlying causal mechanism as a singular one of the APC effects, here we are simply identifying the trend, and acknowledging that it is almost certainly a product of a combination of APC effects.

Whilst this chapter discusses changes over time in a longitudinal framework, this longitudinal element is limited to a single measurement per year. This chapter is essentially sampling in time, with a timestep of a single year, meaning that the variation that occurs within that year cannot be captured by the investigation. Although costly, it would be a very useful future avenue of research to see if there is variability within this one-year time frame and see how that influences the findings of temporal stability within each metric.

5.5.4.3 Modelling Limitations

The sample size used in this investigation was very large for studies of this type, especially for factor analytic studies. Whilst this is the opposite of the common issue for quantitative studies, this does still carry inferential limitation. Effects evaluated by statistical significance are informed by sample size, such that even very small effects may seem significant if enough data is provided to evaluate them. This is not necessarily an issue for this chapter, however the continued improvement of fit with every specification evidenced in Table 5.6 is a product of the sample size. It is thus important to consider that this model is not likely to be the most appropriate for less comprehensive data.

It was not possible to include households in this modelling framework as the Understanding Society data structure generated a new household ID for each and every household at each wave. In a longitudinal framework this proved to be almost exactly collinear with personal identifiers and made model fitting very laborious, as many households have populations of one. Thus, more work needs to be undertaken to understand true household effects, as this structural level was shown to be important in Chapter 4. This would involve the generation of a complex household variable that identified if the same household was being inhabited in sequential waves and accounted for changes in cohabitation across waves.

The final section of results modelled longitudinal trends, with many separately estimated coefficients for each categorical year and demographic variable interaction. This was by design in order to impose as little structure as possible on the model and investigate variability and volatility over the period. However, this did mean that several of the models did not represent a true improvement in fit over the initial fully specified demographic model, due simply to the large number of terms. This is shown in the clear similarities between trends across many of the demographic groups. Thus, it cannot be assumed that for all demographic groups this longitudinal differentiation can objectively be categorised as more than stochastic noise and should be interpreted as such.

6 CONCLUSIONS

6.1.1 Research Outline

This thesis sought to critically and empirically evaluate what is being captured when considering traditional mental health metrics in the UK. Furthermore, it sought to critically evaluate the relationship between mental illness and mental wellbeing through a methodology which allows for interpretive complexity. In doing so, it sought to develop a transferrable method for deconstructing aggregate mental health scores, to allow for greater complexity in the understanding of processes affecting both positive and negative, individual and societal mental health. In light of this greater understanding of what is being inferred from these metrics, it went on to specify complex multilevel models to investigate the differences in mental health across the UK, firstly with the aggregate, cross-level approach to see what could be found with traditional measures, and subsequently with a multivariate, longitudinal approach demonstrating the inference offered by the deconstruction of the GHQ-12. The analysis allowed the investigation of variability between individuals at the geographical, demographic and longitudinal levels, for each of the four differentiated constructs outlined in Chapter 2.

For this final chapter I will give a very brief overview of some of the more interesting findings from the thesis and what adopting this interpretation may offer in terms of future findings. This is followed by a more detailed section outlining the specific contributions of this work, in terms of both methodological and substantive offerings. I will then discuss the policy implications of these findings, followed by a discussion of the key limitations of the work. I will finally discuss directions for future research, some of which is already well under way, and make some concluding remarks.

6.1.2 Summary of Findings

Chapters 2 and 3 addressed the measurement problems that mental health research faces. Chapter 2 challenged the conventional wisdom of treating the GHQ-12 as an additive, unidimensional construct when modelling mental health. Using Exploratory Structural Equation Modelling a model of the GHQ-12 was developed with far superior predictive capacity than the simple summed scored. This involved treating the aggregate GHQ-12 as being underpinned by four separate but related constructs, Lowered Self-Worth, Social Dysfunction, Stress and Emotional Coping. This four-factor solution was demonstrated to be superior to a traditional, unidimensional interpretation both empirically and substantively. Chapter 3 then developed this interpretation, developing what each of these constructs mean in a "wellbeing" sense by evaluating them against constructs derived from a similarly widely used measure to the GHQ-12, the SWE. These findings illustrated remarkable dissimilarity between the wellbeing and mental illness elements, with the GHQ-12 factor scores offering far less predictive capacity in wellbeing responses than initially thought. There were elements of some of the constructs that were related to each other, but there were still notable differences between thematically identical questions between the measures. This reinforces the previous findings as evidence of the need for greater methodological, and syntactic, sensitivity when considering the complex and different processes underpinning mental health.

Chapters 4 and 5 went on to develop explanatory models of mental health across the UK. Chapter 4 took the aggregate responses to the GHQ-12 and SWE from 2009, and specified complex cross-classified multilevel models to investigate which groups were at greatest risk of poor mental health or low wellbeing. This specification highlighted what *could* be found using the unidimensional assumptions of traditional mental health questionnaire literature. There was evidence of considerable differences between the two responses, with worse mental health evidenced for females in the GHQ-12, but no significant differences between sexes in the SWE. Chapter 4 also addressed the geographical determinants of mental illness and wellbeing – finding a greater geography to the latter, and identifying that geography matters most for the youngest and oldest age-groups, with middle-aged individuals tending to experience more homogenous mental health across the UK. Locational versus functional geography was investigated, illustrating that the type of area an individual lived in seemed to matter more for mental health than the specific location, although this result diminished in the face of increased demographic controls, suggesting that the main differences between types of area were largely to do with their demographic make-up. Household was shown to be a critically important spatial scale across both measures. Most clearly though, the results differed strongly between the illness and the wellness measure, reinforcing the call for differentiation between underpinning processes of each from Chapter 3.

Chapter 5 explored the same demographic relationships as Chapter 4 but focused on the GHQ-12 and the constituent elements of it identified in Chapter 2. Specifying each of these as a separate response, and with complex variation allowed between each of the constructs at each spatial scale, we were able to investigate the temporal consistency of these constructs using the same method by which we decomposed variation into area effects in the previous chapter. Of these scalar decompositions, the most important scale was within-individual between-occasions for all but the Lowered Self-Worth response. Stress and Lowered Self-Worth were the most consistent within an individual/household between time points, with Social Dysfunction and Emotional Coping demonstrated to be the most variable within an individual, with the occasion mattering twice as much as the individual in explaining their mental state. There were clear differences between the demographic patterns to each of the four response variables underpinning the GHQ-12. Sex differences were far greater for the Lowered Self-Worth and Stress elements. Sex mattered less for the Social Dysfunction and Emotional Coping constructs which focused on social performance. Similarly, age relationships were markedly different between responses, with the improvement in mental health often advocated by the "Mid-life distress peak" model *only* being evidenced clearly for the Lowered Self-Worth and Stress responses. The years 2010 and 2013 were particularly poor years for mental health across demographics. Differential trends were also modelled for different groups, specifying separate categorical-year terms for each demographic group and illustrating heterogeneity in mental health experience not only between groups, but between responses. Some of the most notable differences were in the experience of the youngest age group, who experienced a large peak in mental distress in the year 2011, and the eldest age group who has a similar peak in 2010. Those of mixed race seemed the most volatile, with blacks the most stable. There was also evidence of a low-stress year in 2012, which was disproportionately enjoyed by whites, males and married individuals.

6.1.3 Methodological Contributions

There are two, key, inter-related methodological contributions in this thesis. Whilst Exploratory Structural Modelling was developed in 2009, its usage has rarely, if ever, extended to big-datasets. It typically remains deployed in educational or psychological studies with sample sizes of under 1000. Similarly, whilst big data studies of mental health as captured by large scale screening questionnaires exist, they rarely go beyond simplistic regression models, and almost never progress beyond summed questionnaire scores for the outcome of interest. The thesis presents methodological advancement on both sides of the equation, developing a complex response interpretation for the dependent variables, and developing analytical complexity for greater explanatory capacity.

Firstly, the outlined ESEM procedure outlined in Chapters 2 and 3 is readily translatable to other mental health responses. This approach is shown to give empirically and substantively better responses to the GHQ-12 and SWE as examples. By explicitly treating the modelled outcomes as imperfect, manifest realisations of true, underpinning processes, ESEM compartmentalises variance into that which is shared between questions, and that which is unique to that question. Using the correlations between these questions, ESEM posits similar processes underpinning the correlated measures. Furthermore, ESEM allows for the relaxation of unrealistic constraints, such as simple structure, necessary for performing traditional factor analysis. Relaxing the assumption of zero cross-loadings allow for the estimation of explicitly correlated factors and does not artificially inflate factor correlations due to conflation of items. Beyond giving improved results, as measured by every goodness-of-fit statistic, the constructs developed here were demonstrated to be *more* substantively dissimilar, due to the specification of non-zero cross loadings giving better attribution of variance. Additionally, we agree with Marsh et al. 2014 in suggesting that this is a substantively intuitive development, as it is highly unlikely that items are informed by a single underpinning process in psychological literature.

The multilevel modelling approach outlined here develops complex notions of geography and gives a quantitative framework for analysing the relative importance of various spatial scales. In specifying complex variability with respect to structural levels, including occasion, we illustrate the capacity to model response interdependence at differing spatial scales. Furthermore, with the specification of cross-classified models including type-of-area as a structural level, I specifically investigated the relative importance of functional versus locational geography in predicting mental health outcomes. Moving beyond geography at the aggregate scale, this was similarly extended to investigating the relative importance of the individual and the occasion. These developments allow the answering of complex questions about the permanence, predictability and ultimately treatment of mental health and illness in the UK. Furthermore, by specifying complex variation with respect to demographic predictors structured by each of these levels, it is possible to investigate for whom geography matters most. This is a key question and allows the investigation to move past positivist assumptions of uniformity within mental health processes for populations.

Finally, combining these two approaches by specifying complex, longitudinal, multivariate, multilevel models affords the benefits of both, without having to succumb to the traditional problems associated with either. The clear difference in interpretive capacity evidenced between Chapters 4 and 5 more clearly demonstrates the importance of properly specified, and rigorously interrogated responses when dealing with a topic as complex as mental health.

6.1.4 Substantive Contributions and Policy Implications

Whilst the methodological contributions are important in the light of mental health research, the overarching substantive contribution is the advocacy and appreciation of complexity in mental health research. The degree of heterogeneity that is uncaptured in traditional interpretations of summed mental health measures is not only clearly non-negligible, but also clearly of substantive interest. I argue here that this nuance in interpretation for mental health is necessary for understanding mental health processes, and indeed for making efficient use of existing resources. It is clearly important to consider differential processes for each response and acknowledge explicitly in the research outline that these processes differ over space and time. Most importantly, it is clear these differences are *not* merely stochastic noise around a true broad pattern that holds across all dimensions of mental health, for all individuals, in all areas, at all times. Rather than any specific technical recommendation, this is the key point to take from this thesis as it places the onus on a shift in the mindset of quantitative researchers when considering mental health, not merely a shift in methodology.

This is in line with an advocacy for complexity in mental health research that has been developing over recent years. This is evidenced both in the innovation of ESEM methods and their relevant application to deepen understanding of composite measures (Marsh *et al.*, 2014), and in the increasing attention paid to positive mental health responses, in whichever form they are considered (Stewart-Brown, Samaraweera, Taggart, N.-B. Kandala, *et al.*, 2015). Whilst these attempts to address the true content and predictors of the GHQ-12 and other mental health measures are undoubtedly important, this must be done with appropriate sensitivity to the complexity of the topic. As such, it seems prudent to stress the importance of methodological sensitivity to the complexity of the topic at the outset of what will presumably be an area of huge innovation in coming years. This means going beyond typically interpreted summed scores of mental health measures whether binary or Likert, and engaging in what is truly underpinning the responses to individual items, with a sensitivity to the notion that this may critically influence what can actually be found. Beyond some notable examples (e.g. Hu *et al.*, 2007), the investigation of the geographical patterning of decomposed composite mental health metrics seems remarkably underexplored in contemporary quantitative geographical literature.

Beyond researchers, this complexity must be acknowledged in policy implementation. Policymakers should be wary of broad-brush approaches to mental health, for there are large, demonstrable differences in mental health between demographic groups, geographical regions and longitudinal timepoints. Policy recommendations should be fully founded, in light of this nuanced interpretation, rather than based on aggregate understandings of mental health. Most critically, as outlined in the latest report of the Chief Medical Officer, research needs to move past the traditional focus on understanding processes of illness and move towards understanding processes of wellness (Davies, 2018), understanding should be sensitive to the different processes governing positive and negative mental health, which were clearly demonstrated in Chapters 3 and 4 . This advocacy of increased complexity should not solely be used to improve forecasting but should also be used to improve the quality of lessons learned from existing literature and data, given the readily extendable methodology.

Beyond this acknowledgement of complexity, there is also evidence of a wide gap between the findings of research centred on self-rated mental health, and that into clinical outcomes such as suicide. Different risk-groups are identified when considering each of these responses. Whilst this is clearly plausible, there is a very real difficulty in isolating individuals' reluctance to admit feeling negative, from actively feeling positive, these are manifested identically in mental health questionnaire response, and there is little that can be done methodologically to improve upon this. Whilst the Emotional Coping construct was suggested to perhaps shed some light on these differences, the trends that were identified within it were not ultimately similar to those in suicide literature. Whilst the difference between males and females was reduced relative to the other constructs, it was not reversed. Similarly, there was not a large divergence between the trends of different genders on any of the constructs mirroring the disproportionate increase in male suicides over recent years (Barr *et al.*, 2012). Whilst there is some suggestion here of the differences between stoic and non-stoic groups being more easily evidenced in positive mental health metrics, it is still statistically impossible to separate the two conceptually. The key policy implementation here then is advocacy of greater destigmatisation and awareness of mental health far more broadly. Given the social comparison elements of mental distress, the more awareness people have of the concept of mental health, the more likely people are to be accurate in characterising their mental health. This accuracy in reporting will foster greater accuracy in population-screening, ultimately leading to better recommendations for policy. We are already seeing efforts to destigmatise mental health, and whilst these are usually framed in terms of benefit for the individual, they potentially carry the unforeseen benefit of improved mental health at the aggregate level through better monitoring and forecasting, as well as the obvious benefits of a society where an individual did not feel uncomfortable speaking out about their distress, whatever the manifestation.

6.1.5 Key Limitations

Issues outlined here are undoubtedly important, however it is not the intended suggestion that the models presented here are the end point of the long-run development of statistical complexity in modelling. These limitations are true for the models presented, but the precise aim of modelling is to simplify reality to a point where we can garner useful insight into it. Mental health research should be guided, principally, by research questions and available data, the limitations of models should be fully and explicitly addressed, but do not render the findings of those models redundant. We do however advocate for careful considerations of the shortcomings of these models and their associated results. The shortcomings of chapter-specific methodologies have been discussed at length in respective chapters, so here we present some key theoretical limitations of inferring to wider populations about mental health experience.

Firstly, in all models presented, there is the capacity for the biasing of results by unobserved social processes or variables. As outlined at several points in this thesis, the largest issue for any self-rated mental health investigation is the conflation of reluctance to respond negatively with positive mental health experience. This is the key limitation in any investigation of self-rated mental health. All recommendations are potentially biased towards those individuals capable of accurately identifying their mental health status. Similarly, conclusions possibly but unknowably understate the mental distress of those individuals who, consciously or otherwise, are less capable of expressing negative emotion. This could conceivably function in both directions, understating both positive and negative emotion for the stoic and biasing results towards the expressive. There is no simple statistical fix for this problem. This was observed in the Emotional Coping metric which initially looked as though it may offer insight into the differentially stoic response tendencies of demographic groups. However, fundamentally it still required individuals to acknowledge feeling negative, but think that it didn't affect their functioning. As such, it still succumbs to the same bias as all self-reporting metrics, in that it cannot capture individuals who are truly stoic, and therefore don't acknowledge negative emotion in their responses in the first place.

Unobserved variables and processes are likely to also bias causal inference. It is common for policy-makers or lay-audiences to readily leap to causal arguments when presented with a finding that females experience worse mental health than males. Whilst it is useful to know that females on average experience worse mental health than men it does not necessarily imply there is something *inherent* about being a woman that is unhealthy. Common wisdom too readily simplifies this to assume that females are invariably and innately less happy than males, however this could very plausibly be the lived experience of being a woman. Social phenomena such as stigma, discrimination, wage gaps and barriers to confidence are patterned demographically and exist at a structural level in-built to the lived experience of individuals. These factors clearly influence mental health, and this potential complexity should be acknowledged in the all-too-often oversimplified interpretation of causality.

Finally, this thesis has generalised at several points from questionnaire metrics to broad mental states, often using these arguments to subsequently recommend against doing exactly that. Mental wellbeing and mental illness are clearly more heterogeneous than is commonly thought or operationalised in quantitative research, and the over-arching theme of this thesis is to caution against doing that. Whilst it has been useful to demonstrate the dissimilarity of the processes underpinning measures generally considered to measure wellbeing and illness, this should not be taken as evidence that they are unrelated in all contexts and at all times. It has empirically gone no further than demonstrating dissimilarity between the SWE and the GHQ-

12. This should be carefully considered when generalising to policy, as if nothing else this thesis has demonstrated the necessity of nuance in drawing conclusions.

6.1.6 Future Study

Continuing use of these constructs in mental health modelling is something we advocate here. However, given the acknowledged complexity of the topic, this understanding would require regular re-evaluation. In light of greater methodological capacity, we do not imagine that this would be the end of the development of mental health investigation in the UK. There will be further developments in coming years, and there are already campaigns for mental health promotion nationwide for at-risk groups which could be evaluated using this methodology. It would be fascinating to model the changing differences between sexes for the responses in the light of increased mental health awareness among those groups which are traditionally the most stoic. As outlined here, there is no real gauge for the degree of stoicism within an individual, and the degree to which they are likely to respond honestly to a mental health questionnaire.

Chapter 5 alluded to the possibility of extending the multilevel framework to include structural levels that were present in Chapter 4. These were not carried out due to time constraints, but it would be of clear benefit to see if similarities within households, or within types of areas, were increasing or decreasing over time, and whether this was consistent for all demographic groups. We highly recommend this further investigation be carried out to better characterise the rapidly changing area of mental health geography.

Whilst this thesis was concerned solely with the GHQ-12 and the SWE as responses, the methodology is intentionally generalisable. The same procedure could be used for any multi-item instrument, and researchers are encouraged to do so. Including other mental health outcomes would improve the understanding of underpinning processes outlined here and could offer insight into the true nature of underpinning constructs within these instruments. Specifically modelled factor correlations between underpinning constructs of different mental health questionnaires would provide valuable information about the nature of, and processes underpinning, wellbeing, mental illness and mental health more broadly.

Finally, it was also outlined early on in this thesis that there has been a wealth of data already collected on the GHQ-12. The purpose of highlighting the shortcomings of the GHQ-12 as a screening instrument for mental health was not to suggest the discarding of decades of research. The methods proposed here offer insight into how this information could be reinterpreted in light of greater methodological understanding. This could and should be used to contribute to the emerging newer understandings of mental health.

6.1.7 Concluding Remarks

It is simple and easy to assume that different questions pertaining to mental health, all referring to some sort of psychological distress are all capturing the same underpinning dimension of “sadness”, or even “absence of happiness” in the case of some metrics. The idea being that whilst all questions on mental health screening questionnaires might receive *subtly* different responses, we can view that as stochastic variation, and treat them as flawed realisations of the same process. This view of mental health is simple to operationalise, and simple to make recommendations for. The only issue is that it is demonstrably false. Mental health is complex, we have seen the beginnings of this understanding with the two-continua model of mental health, and the acknowledgement of the differences between the determinants of positive and negative mental health. We have gone further here and established that even within globally recognised and incredibly widely used and accepted mental health metrics, there is great variation, and that perhaps this large-scale positivist approach of understanding mental health is at best outdated, and at worst actively misleading. Where individuals live is differentially important for different groups, certain regions have disproportionately worse mental health, different demographic groups illustrate very different trends, positive and negative mental health have different determinants, different demographic patterning, different trajectories and different geographies. Different constructs within the *same*, well validated, globally used instrument have wildly different patterning.

All of these arguments point to the inherent complexity of mental health as a construct. For us to make robust, reasonable inference, this conceptual complexity *needs* to be met with an adequate modelling framework that allows for the estimation of these differences, or we risk misattributing problems, or even worsening mental health for all but the statistically average respondent. This thesis clearly outlines a transferrable procedure for addressing and quantifying to some extent the conceptual complexity present in mental health responses, which we suggest is used in future discussion of mental health evaluation. Furthermore, it raises important questions about the nature of nationwide screening for something as nuanced as mental health, and even about the nature and meaning of *illness* in a broader setting.

Finally, as evidenced in the complicated discussion sections in each chapter, it also addresses the need for greater awareness of mental health nationwide. Issues of stigma and awareness of mental health have been demonstrated repeatedly to influence individual behaviour in usage of mental health services (Wang *et al.*, 2007; Henderson, Evans-Lacko and Thornicroft, 2013; Clement *et al.*, 2015; Hunt *et al.*, 2015; Angermeyer *et al.*, 2017). However, this thesis demonstrates that beyond the obvious individual benefits of greater mental health awareness in help-seeking, benefits of awareness go beyond the individual. If individuals are more aware of their mental health and stigma is reduced, then those individuals are likely to be more capable of discussing it. This would be directly beneficial to the population as it would allow researchers to more accurately model mental distress and target policy and help to where it is most needed. Hopefully this goes may go some way to addressing the existing gaps between the conclusions of investigations of self-rated mental health and clinical outcomes such as suicide. It seems that beyond individuals, populations are healthier when they are talking about mental health.

7 REFERENCES

Aguado, J. *et al.* (2012) 'Examining the factor structure and discriminant validity of the 12-item General Health Questionnaire (GHQ-12) among Spanish postpartum women.', *Assessment*, 19(4), pp. 517–25. doi: 10.1177/1073191110388146.

Anastasi, A. (1976) *Psychological testing, 4th ed.*, *Psychological testing, 4th ed.*

Andrich, D. and Schoubroeck, L. Van (1989) 'The General Health Questionnaire: a psychometric analysis using latent trait theory', *Psychological Medicine*, 19(2), pp. 469–485. doi: 10.1017/S0033291700012502.

Angermeyer, M. C. *et al.* (2017) 'Public attitudes towards psychiatry and psychiatric treatment at the beginning of the 21st century: a systematic review and meta-analysis of population surveys', *World Psychiatry*, 16(1), pp. 50–61. doi: 10.1002/wps.20383.

Arcaya, M. *et al.* (2012) 'Area variations in health: A spatial multilevel modeling approach', *Health and Place*, 18(4), pp. 824–831. doi: 10.1016/j.healthplace.2012.03.010.

Arokiasamy, P. *et al.* (2015) 'The impact of multimorbidity on adult physical and mental health in low- and middle-income countries: What does the study on global ageing and adult health (SAGE) reveal?', *BMC Medicine*. BMC Medicine, 13(1), pp. 1–16. doi: 10.1186/s12916-015-0402-8.

Arpino, B. and Varriale, R. (2010) *Assessing the quality of institutions' rankings obtained through multilevel linear regression models.*, *Carlo F. Dondena Centre for Reserach on Social Dynamics*. doi: 10.1016/B978-0-12-417113-8.00008-4.

Artazcoz, L. *et al.* (2004) 'Unemployment and mental health: understanding the interactions among gender, family roles, and social class', *American Journal of Public Health*, 94(1), pp. 82–88. doi: 10.2105/AJPH.94.1.82.

Asparouhov, T. and Muthen, B. (2009) *Exploratory Structural Equation Modeling*, *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705510903008204.

Asparouhov, T. and Muthén, B. (2009) *Exploratory Structural Equation Modeling*, *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705510903008204.

Asparouhov, T. and Muthén, B. (2010) 'Bayesian analysis using Mplus: Technical implementation', *Los Angeles: Muthén & Muthén*, pp. 1–38.

Asparouhov, T., Muthén, B. and Morin, A. J. S. (2015) 'Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeyer et al. ', *Journal of Management* , 41(6), pp. 1561–1577. doi: 10.1177/0149206315591075.

Asparouhov, T. and Muthén, B. O. (2009) *Exploratory Structural Equation Modeling*, *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705510903008204.

Assari, S., Watkins, D. C. and Caldwell, C. H. (2015) 'Race Attribution Modifies the

Association Between Daily Discrimination and Major Depressive Disorder Among Blacks: the Role of Gender and Ethnicity', *Journal of racial and ethnic health disparities*, 2(2), pp. 200–210. doi: 10.1007/s40615-014-0064-9.

Baker, E., Bentley, R. and Mason, K. (2013) 'The Mental Health Effects of Housing Tenure: Causal or Compositional?', *Urban Studies*, 50(2), pp. 426–442. doi: 10.1177/0042098012446992.

Baksheev, G. N. *et al.* (2011) 'Validity of the 12-item General Health Questionnaire (GHQ-12) in detecting depressive and anxiety disorders among high school students.', *Psychiatry research*. Elsevier Ireland Ltd, 187(1–2), pp. 291–6. doi: 10.1016/j.psychres.2010.10.010.

Barr, B. *et al.* (2012) 'Suicides associated with the 2008-10 economic recession in England: time trend analysis', *Bmj*, 345(aug13 2), pp. e5142–e5142. doi: 10.1136/bmj.e5142.

Barr, B., Kinderman, P. and Whitehead, M. (2015) 'Trends in mental health inequalities in England during a period of recession, austerity and welfare reform 2004 to 2013', *Social Science and Medicine*. Elsevier Ltd, 147, pp. 324–331. doi: 10.1016/j.socscimed.2015.11.009.

Barry, M. M. (2009) 'Addressing the Determinants of Positive Mental Health: Concepts, Evidence and Practice', *International Journal of Mental Health Promotion*, 11(3), pp. 4–17. doi: 10.1080/14623730.2009.9721788.

Bartram, D. J., Sinclair, J. M. and Baldwin, D. S. (2013) 'Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis.', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 22(2), pp. 379–91. doi: 10.1007/s11136-012-0144-4.

Beddington, J. *et al.* (2008) 'The mental wealth of nations.', *Nature*, 455(7216), pp. 1057–1060. doi: 10.1038/4551057a.

Bell, A. (2014) 'Life-course and cohort trajectories of mental health in the UK, 1991-2008 - A multilevel age-period-cohort analysis', *Social Science and Medicine*. Elsevier Ltd, 120, pp. 21–30. doi: 10.1016/j.socscimed.2014.09.008.

Bell, A., Fairbrother, M. and Jones, K. (2016) 'Fixed and Random effects : making an informed choice', (March), pp. 1–38.

Bell, A. and Jones, K. (2013) 'The impossibility of separating age, period and cohort effects', *Social Science & Medicine*, 93, pp. 163–165. doi: 10.1016/j.socscimed.2013.04.029.

Bell, A. and Jones, K. (2018) 'The hierarchical age–period–cohort model: Why does it find the results that it finds?', *Quality and Quantity*. Springer Netherlands, 52(2), pp. 783–799. doi: 10.1007/s11135-017-0488-5.

Bentler, P. M. (2007) 'On tests and indices for evaluating structural models', *Personality and Individual Differences*, 42(5), pp. 825–829. doi: 10.1016/j.paid.2006.09.024.

Bentley, R. *et al.* (2011) 'Association between housing affordability and mental health: A longitudinal analysis of a nationally representative household survey in Australia', *American Journal of Epidemiology*, 174(7), pp. 753–760. doi: 10.1093/aje/kwr161.

van den Berg, A. E. *et al.* (2010) 'Green space as a buffer between stressful life events and health.', *Social science & medicine (1982)*. Elsevier Ltd, 70(8), pp. 1203–10. doi: 10.1016/j.socscimed.2010.01.002.

Berg, A. O. *et al.* (2011) 'Perceived discrimination is associated with severity of positive and depression/anxiety symptoms in immigrants with psychosis: A cross-sectional study', *BMC Psychiatry*, 11. doi: 10.1186/1471-244X-11-77.

Berkey, L., Frey, B. S. and Stutzer, A. (2003) 'Happiness and Economics: How the Economy and Institutions Affect Human Well-Being', *Contemporary Sociology*, 32(2), p. 190. doi: 10.2307/3089588.

Bhui, K. *et al.* (2005) 'Racial/ethnic discrimination and common mental disorders among workers: Findings from the EMPIRIC study of ethnic minority groups in the United Kingdom', *American Journal of Public Health*, 95(3), pp. 496–501. doi: 10.2105/AJPH.2003.033274.

Blanchflower, D. G. and Oswald, A. J. (2008) 'Is well-being U-shaped over the life cycle?', *Social science & medicine (1982)*, 66(8), pp. 1733–49. doi: 10.1016/j.socscimed.2008.01.030.

Bleil, M. E. *et al.* (2008) 'Trait negative affect: toward an integrated model of understanding psychological risk for impairment in cardiac autonomic function.', *Psychosomatic medicine*, 70(3), pp. 328–37. doi: 10.1097/PSY.0b013e31816baefa.

Bloom, D. E. *et al.* (2011) *The Global Economic Burden of Noncommunicable Diseases*, *World Economic Forum*. Available at: http://ideas.repec.org/p/gdm/wpaper/8712.html.

Boehm, J. *et al.* (2012) 'A Prospective Study of Positive Psychological Well-Being and Coronary Heart Disease', 30(3), pp. 259–267. doi: 10.1037/a0023124.A.

Boehm, J. K. and Kubzansky, L. D. (2012) 'The heart's content: The association between positive psychological well-being and cardiovascular health', *Psychological Bulletin*, 138(4), pp. 655–691. doi: 10.1037/a0027448.

Borgers, N., Hox, J. and Sikkel, D. (2004) 'Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point', *Quality and Quantity*, 38(1), pp. 17–33. doi: 10.1023/B:QUQU.0000013236.29205.a6.

Bowling, A. (2005) 'Mode of questionnaire administration can have serious effects on data quality', *Journal of Public Health*, 27(3), pp. 281–291. doi: 10.1093/pubmed/fdi031.

Boyd, A. *et al.* (2013) 'Cohort profile: The "Children of the 90s"-The index offspring of the avon longitudinal study of parents and children', *International Journal of Epidemiology*, 42(1), pp. 111–127. doi: 10.1093/ije/dys064.

Boyd, R. C., Le, H. N. and Somberg, R. (2005) 'Review of screening instruments for postpartum depression', *Archives of Women's Mental Health*, 8(3), pp. 141–153. doi: 10.1007/s00737-005-0096-6.

Bracke, P., Pattyn, E. and Von dem Knesebeck, O. (2013) 'Overeducation and depressive symptoms: Diminishing mental health returns to education', *Sociology of Health and Illness*, 35(8), pp. 1242–1259. doi: 10.1111/1467-9566.12039.

Bracke, P., Van De Straat, V. and Missinne, S. (2014) 'Education, mental health, and education–labor market misfit', *Journal of Health and Social Behavior*, 55(4), pp. 442–459. doi: 10.1177/0022146514557332.

Bradburn Norman, M. (1964) 'The Structure of Psychological Well-Being', *Physica Status Solidi B*, pp. 273–277. doi: 10.1002/pssb.19640050205.

Braun, A. A. (2009) 'Gross National Happiness in Bhutan: A Living Example of an Alternative Approach to Progress', *Financial Times*, (September 24). Available at: http://www.grossnationalhappiness.com/OtherArticles/GNHPaperbyAlejandro.pdf.

Breier, a *et al.* (1991) 'National Institute of Mental Health longitudinal study of chronic schizophrenia. Prognosis and predictors of outcome.', *Archives of general psychiatry*, 48(3), pp. 239–246. doi: 10.1001/archpsyc.1991.01810270051007.

Brodersen, J. *et al.* (2007) 'Measuring the psychosocial consequences of screening', *Health and Quality of Life Outcomes*, 5(1), p. 3. doi: 10.1186/1477-7525-5-3.

Browne, W. J. and Draper, D. (2006) 'A comparison of Bayesian and likelihood-based methods for fitting multilevel models', (3), pp. 473–514.

Bryman, A. (1988) *Quantity and quality in social research*, *Contemporary Social Research Series, (8)*.

Burton-jeangros, C. and Editors, D. B. (2015) 'A Life Course Perspective on Health Trajectories and Transitions', *Springer*, 4, pp. 39–61. doi: 10.1007/978-3-319-20484-0.

Cacioppo, J. T. and Berntson, G. G. (1999) 'The affect system: Architecture and operating characteristics', *Current Directions in Psychological Science*, 8(5), pp. 133–137. doi: 10.1111/1467-8721.00031.

Cairney, J. and Krause, N. (2005) 'The social distribution of psychological distress and depression in older adults.', *Journal of aging and health*, 17(6), pp. 807–35. doi: 10.1177/0898264305280985.

Campbell, A. and Knowles, S. (2007) 'A confirmatory factor analysis of the GHQ12 using a large Australian sample', *European Journal of Psychological Assessment*, 23(1), pp. 2–8. doi: 10.1027/1015-5759.23.1.2.

Campbell, A., Walker, J. and Farrell, G. (2003) 'Confirmatory factor analysis of the GHQ-12: can I see that again?', *Australian and New Zealand Journal of Psychiatry*, 37(4), pp. 475–483. doi: 10.1046/j.1440-1614.2003.01208.x.

Castellví, P. *et al.* (2014) 'The Spanish version of the Warwick-Edinburgh mental well-being scale (WEMWBS) is valid for use in the general population.', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 23(3), pp. 857–68. doi: 10.1007/s11136-013-0513-7.

Chaix, B. *et al.* (2005) 'Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malm??, Sweden, 2001', *American Journal of Epidemiology*, 162(2), pp. 171–182. doi: 10.1093/aje/kwi175.

Chambers, J. M. *et al.* (1983) *Graphical Methods for Data Analysis, Wadsworth International Group Belmont*.

Chanfreau, J. *et al.* (2013) 'Predicting wellbeing', p. 150. Available at: http://www.natcen.ac.uk/study/predictors-of-wellbeing.

Chen, F. F. (2007) 'Sensitivity of goodness of fit indexes to lack of measurement invariance', *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), pp. 464–504. doi: 10.1080/10705510701301834.

Cheung, Y. B. (2002) 'A confirmatory factor analysis of the 12-item General Health Questionnaire among older people', (April), pp. 739–744.

Chevalier, A. and Feinstein, L. (2006) 'Sheepskin or Prozac: The causal effect of education on mental health', (2231). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=923530 (Accessed: 24 January 2013).

Chida, Y. and Steptoe, A. (2008) 'Positive psychological well-being and mortality: a quantitative review of prospective observational studies.', *Psychosomatic medicine*, 70(7), pp. 741–56. doi: 10.1097/PSY.0b013e31818105ba.

Clapham, D., Foye, C. and Christian, J. (2017) 'The Concept of Subjective Well-being in Housing Research', *Housing, Theory and Society*. Routledge, 6096, pp. 1–20. doi: 10.1080/14036096.2017.1348391.

Clark, A. E. and Oswald, A. J. (2011) 'The curved relationship between subjective well-being and age To cite this version ':

Clark, L. A., Watson, D. and Mineka, S. (1994) 'Temperament, Personality, and the Mood and Anxiety Disorders', *Journal of Abnormal Psychology*, pp. 103–116. doi: 10.1037/0021-843X.103.1.103.

Clarke, A. *et al.* (2011) 'Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment.', *BMC public health*. BioMed Central Ltd, 11(1), p. 487. doi: 10.1186/1471-2458-11-487.

Clement, S. *et al.* (2015) 'What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies', *Psychological Medicine*, 45(1), pp. 11–27. doi: 10.1017/S0033291714000129.

Collins, D. L., Baum, A. and Singer, J. E. (1983) 'Coping with chronic stress at Three Mile Island: Psychological and biochemical evidence.', *Health Psychology*, 2(2), pp. 149–166. doi: 10.1037/0278-6133.2.2.149.

Collishaw, S. (2015) 'Annual research review: Secular trends in child and adolescent mental health', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56(3), pp. 370–393. doi: 10.1111/jcpp.12372.

Connell, J., O'Cathain, A. and Brazier, J. (2014) 'Measuring quality of life in mental health: are we asking the right questions?', *Social science & medicine (1982)*. Elsevier Ltd, 120, pp. 12–20. doi: 10.1016/j.socscimed.2014.08.026.

Corrigan, P. W. *et al.* (2012) 'Challenging the Public Stigma of Mental Illness: A Meta-Analysis of Outcome Studies', *Psychiatric Services*, 63(10), pp. 963–973. doi: 10.1176/appi.ps.201100529.

Cosco, T. D., Howse, K. and Brayne, C. (2017) 'Healthy ageing, resilience and wellbeing', *Epidemiology and Psychiatric Sciences*, (2017), pp. 1–5. doi: 10.1017/S2045796017000324.

Creed, P. A. and Evans, B. M. (2002) 'Personality, well-being and deprivation theory', *Personality and Individual Differences*, 33(7), pp. 1045–1054. doi: 10.1016/S0191-8869(01)00210-0.

Cruise, S. M. and O'Reilly, D. (2015) 'Are the differences in adulthood ill-health across the north-south divide and between Scotland and England also evident in early childhood health indicators?', *Social Science and Medicine*. Elsevier Ltd, 130, pp. 277–283. doi: 10.1016/j.socscimed.2015.02.021.

Daniel, E. and Bright, G. (2011) 'Exploring the geographical distribution of wealth using the output area classification', *Economic & Labour Market Review*, 5(1), pp. 59–78. doi: 10.1057/elmr.2011.7.

Davies, S. (2013) 'Annual Report of the Chief Medical Officer 2013, Public Health Mental Priorities: Investing in the evidence', p. 320. Available at: https://www.gov.uk/government/organisations/department-of-health.

Davies, S. C. (2014) 'Annual Report of the Chief Medical Officer 2013. Public Mental Health Priorities: Investing in the Evidence', *London: Department of Health*. doi: 10.12968/nuwa.2009.24.3.1092615.

Davies, S. C. (2018) 'Annual Report of the Chief Medical Officer, 2018: Health 2040 - Better Health Within Reach'.

Department of Health (2013) 'Including a measure of mental wellbeing in the ONS Wellbeing Framework - Measuring National Wellbeing Technical Advisory Group Paper', (April), pp. 2–7.

Department of Health, M. H. D. (2009) 'New Horizons: Towards a shared vision for mental health.', pp. 3–5.

Department of Health and Health, D. of (2011) *No health without mental health, Delivering better mental health outcomes for people of all ages, Statistics*.

Diener, E. and Chan, M. Y. (2011) 'Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity', *Applied Psychology: Health and Well-Being*, 3(1), pp. 1–43. doi: 10.1111/j.1758-0854.2010.01045.x.

Diez-Roux, A. V. (1998) 'Bringing context back into epidemiology: Variables and fallacies in multilevel analysis', *American Journal of Public Health*, 88(2), pp. 216–222. doi: 10.2105/AJPH.88.2.216.

Doi, Y. and Minowa, M. (2003) 'Factor structure of the 12- item General Health Questionnaire in the Japanese general adult population', *Psychiatry and clinical neurosciences*, 57(4), pp. 379–383. doi: 10.1046/j.1323-1316.2003.01135.x.

Dolan, P. and Peasgood, T. (2006) 'Review of research on the influences on personal well-being and application to policy making', *Final report for Defra*, (August).

Eack, S. M. and Newhill, C. E. (2007) 'Psychiatric symptoms and quality of life in schizophrenia: A meta-analysis', *Schizophrenia Bulletin*, 33(5), pp. 1225–1237. doi: 10.1093/schbul/sbl071.

Easterlin, R. a *et al.* (2010) 'The happiness-income paradox revisited.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(52), pp. 22463–8. doi: 10.1073/pnas.1015962107.

Ebrahim, S. and Smith, G. D. (2013) 'Commentary: Should we always deliberately be non-representative?', *International Journal of Epidemiology*, 42(4), pp. 1022–1026. doi: 10.1093/ije/dyt105.

Evans, G. W. (2003) 'The built environment and mental health.', *Journal of urban health : bulletin of the New York Academy of Medicine*, 80(4), pp. 536–55. doi: 10.1093/jurban/jtg063.

Ezzy, D. (1993) 'UNEMPLOYMENT AND MENTAL HEALTH ':, 37(I), pp. 41–52.

Fabrigar, L. R. *et al.* (1999) 'Evaluating the use of exploratory factor analysis in psychological research.', *Psychological Methods*, 4(3), pp. 272–299. doi: 10.1037//1082-989X.4.3.272.

Faris, R. E. L. and Dunham, H. W. (1939) *Mental disorders in urban areas: an ecological study of schizophrenia and other psychoses.* Oxford, England: Univ. Chicago Press.

Fergusson, D. M. and Woodward, L. J. (2002) 'Mental Health, Educational, and Social Role Outcomes of Adolescents With Depression', *Archives of General Psychiatry*, 59(3), p. 225. doi: 10.1001/archpsyc.59.3.225.

Fernandes, H. M. and Vasconcelos-Raposo, J. (2013) 'Factorial validity and invariance of the GHQ-12 among clinical and nonclinical samples.', *Assessment*, 20(2), pp. 219–29. doi: 10.1177/1073191112465768.

Ferrari, A. J. *et al.* (2013) 'Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010', *PLoS Medicine*, 10(11). doi: 10.1371/journal.pmed.1001547.

Ferrie, J. E. (2001) 'Is job insecurity harmful to health?', *Journal of the Royal Society of Medicine*, 94(2), pp. 71–76. doi: 10.1177/014107680109400206.

Fioravanti, M. *et al.* (2005) 'A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia', *Neuropsychology Review*, 15(2), pp. 73–95. doi: 10.1007/s11065-005-6254-9.

Flint, E., Bartley, M., *et al.* (2013) 'Do labour market status transitions predict changes in psychological well-being?', *Journal of Epidemiology and Community Health*, 67(9), pp. 796–802. doi: 10.1136/jech-2013-202425.

Flint, E., Shelton, N., *et al.* (2013) 'Do local unemployment rates modify the effect of individual labour market status on psychological distress?', *Health and Place*. Elsevier, 23, pp. 1–8. doi: 10.1016/j.healthplace.2013.04.004.

Folkman, S. *et al.* (1986) 'Appraisal, coping, health status, and psychological symptoms.', *Journal of Personality and Social Psychology*, 50(3), pp. 571–579. doi: 10.1037/0022-3514.50.3.571.

Fone, D. *et al.* (2013) 'Common mental disorders, neighbourhood income inequality and income deprivation: Small-area multilevel analysis', *British Journal of Psychiatry*, 202(4). doi: 10.1192/bjp.bp.112.116178.

Frech, A. and Williams, K. (2007) 'Depression and the psychological benefits of entering marriage', *Journal of Health and Social Behavior*, 48(2), pp. 149–163. doi: 10.1177/002214650704800204.

French, D. J. and Tait, R. J. (2004) 'Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents.', *European child & adolescent psychiatry*, 13(1), pp. 1–7. doi: 10.1007/s00787-004-0345-7.

Fryers, T. *et al.* (2005) 'The distribution of the common mental disorders: social inequalities in Europe.', *Clinical practice and epidemiology in mental health : CP & EMH*, 1, p. 14. doi: 10.1186/1745-0179-1-14.

Galea, S. *et al.* (2005) 'Urban built environment and depression: a multilevel analysis.', *Journal of epidemiology and community health*, 59(10), pp. 822–7. doi: 10.1136/jech.2005.033084.

Gao, F. *et al.* (2004) 'Does the 12-item General Health Questionnaire contain multiple factors and do we need them?', *Health and quality of life outcomes*, 2, p. 63. doi: 10.1186/1477-7525-2-63.

Gareth, H. and Wilson, C. (2018) 'From transcendence to general maintenance: Exploring the creativity and wellbeing dynamic in higher education.', in Reisman, F. (ed.) *2018 International Conference on Knowledge, Innovation and Enterprise*. London: KIE Conference Publications. Availa-.

Garrard, J. *et al.* (1998) 'Clinical detection of depression among community-based elderly people with self-reported symptoms of depression.', *The Journals Of Gerontology. Series A, Biological Sciences And Medical Sciences*, 53(2), pp. M92–M101.

Gelman, a *et al.* (1996) 'Posterior predictive assessment of model fitness via realized discrepancies. Vol.6, No.4.', *Statistica Sinica*, 6(4), pp. 733–807. doi: 10.1.1.142.9951.

Gibson, M. *et al.* (2011) 'Housing and health inequalities: A synthesis of systematic reviews of interventions aimed at different pathways linking housing and health', *Health and Place*. Elsevier, 17(1), pp. 175–184. doi: 10.1016/j.healthplace.2010.09.011.

Giggs, J. (1973) 'The distribution of schizophrenics in Nottingham', *Transactions of the Institute of British Geographers*, 59(59), pp. 55–76. Available at: http://www.jstor.org/stable/10.2307/621712 (Accessed: 2 November 2012).

Goldberg, D. *et al.* (1997) 'The validity of two versions of the GHQ in the WHO study of mental illness in general health care', *Psychological Medicine*, 27(1), pp. 191–197. Available at: http://journals.cambridge.org/abstract_S0033291796004242 (Accessed: 20 January 2013).

Goldberg, D. P. (1972) *The detection of psychiatric illness by questionnaire: A technique for the identification and assessment of non-psychotic psychiatric illness.* Oxford, England: Oxford U. Press.

Goldberg, D. P. and Hillier, V. F. (1979) 'A scaled version of the General Health Questionnaire.', *Psychological medicine*, 9, pp. 139–145. doi: 10.1017/S0033291700021644.

Goldberg, D. P., Oldehinkel, T. and Ormel, J. (1998) 'Why GHQ threshold varies from one place to another.', *Psychological medicine*, 28(4), pp. 915–21. doi: 10.1017/S0033291798006874.

Goldstein, H. *et al.* (2015) 'Population sampling in longitudinal suverys', *Longitudinal and Life Course Studies*, 6(4), pp. 447–475. doi: 10.14301/llcs.v6i4.345.

Goldstein, H., Browne, W. and Rasbash, J. (2002) 'Multilevel modelling of medical data', *Statistics in Medicine*, 21(21), pp. 3291–3315. doi: 10.1002/sim.1264.

Goodchild, M. E. and Duncan-Jones, P. (1985) 'Chronicity and the General Health Questionnaire', *The British Journal of Psychiatry*, 146(1), pp. 55–61. doi: 10.1192/bjp.146.1.55.

Gouveia, V. V *et al.* (2010) 'Factorial validity and reliability of the General Health Questionnaire (GHQ-12) in the Brazilian physician population.', *Cadernos de Saúde Pública*, 26(7), pp. 1439–1445. doi: 10.1590/S0102-311X2010000700023.

Gove, W. R., Style, C. B. and Hughes, M. (1990) 'The Effect of Marriage on the Well-Being of Adults: A Theoretical Analysis', *Journal of Family Issues*, 11(1), pp. 4–35. doi: 10.1177/019251390011001002.

Graetz, B. (1991) 'Multidimensional properties of the general health questionnaire', *Social psychiatry and psychiatric epidemiology*, pp. 132–138. Available at: http://link.springer.com/article/10.1007/BF00782952 (Accessed: 16 February 2015).

Grant, F., Guille, C. and Sen, S. (2013) 'Well-Being and the Risk of Depression under Stress', *PLoS ONE*, 8(7), pp. 1–6. doi: 10.1371/journal.pone.0067395.

Greenspoon, P. J. and Saklofske, D. H. (2001) 'Toward an integration of subjective well-being and psychopathology', *Social Indicators Research*, 54(1), pp. 81–108. doi: 10.1023/A:1007219227883.

Grundy, E. and Sloggett, A. (2003) 'Health inequalities in the older population: the role of personal capital, social resources and socio-economic circumstances.', *Social science & medicine (1982)*, 56(5), pp. 935–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12593868.

Häfner, H. (2003) 'Gender differences in schizophrenia', *Psychoneuroendocrinology*, 28, pp. 17–54. doi: 10.1016/S0306-4530(02)00125-7.

Halpern, D. (1993) 'Minorities and mental health', *Social Science and Medicine*, 36(5), pp. 597–607. doi: 10.1016/0277-9536(93)90056-A.

Hankins, M. (2008a) 'Clinical Practice and Epidemiology The factor structure of the twelve item General Health Questionnaire (GHQ-12): the result of negative phrasing ?', 8, pp. 1–8. doi: 10.1186/1745-0179-4-Received.

Hankins, M. (2008b) 'The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions.', *BMC public health*, 8, p. 355. doi: 10.1186/1471-2458-8-355.

Harding, C. M. *et al.* (1987) 'Chronicity in schizophrenia: fact, partial fact, or artifact?', *Hosp.Community Psychiatry*, 38(5), pp. 477–486. doi: 10.1176/ps.38.5.477.

Harris, R., Johnston, R. and Burgess, S. (2007) 'Neighborhoods, ethnicity and school choice: Developing a statistical framework for geodemographic analysis', *Population Research and Policy Review*, 26(5–6), pp. 553–579. doi: 10.1007/s11113-007-9042-9.

Harrison, G. *et al.* (2009) 'A prospective study of severe mental disorder in Afro-Caribbean patients', *Psychological Medicine*, 18(3), p. 643. doi: 10.1017/S0033291700008321.

Headey, B., Kelley, J. and Wearing, a (1993) 'Dimensions of mental health: life satisfaction, positive affect, anxiety and depression', *Social Indicators Research*, pp. 63–82. doi: 10.1007/BF01136197.

Helmchen, L. a and Lo Sasso, A. T. (2010) 'How sensitive is physician performance to alternative compensation schedules? Evidence from a large network of primary care clinics.', *Health economics*, 19(11), pp. 1300–1317. doi: 10.1002/hec.

Henderson, C., Evans-Lacko, S. and Thornicroft, G. (2013) 'Mental illness stigma, help seeking, and public health programs', *American Journal of Public Health*, 103(5), pp. 777–780. doi: 10.2105/AJPH.2012.301056.

Henderson, S. (1981) 'Social relationships, adversity and neurosis: An analysis of prospective observations', *British Journal of Psychiatry*, 138(5), pp. 391–398. doi: 10.1192/bjp.138.5.391.

Hiscock, R. *et al.* (2003) 'Residents and residence: Factors predicting the health disadvantage of social renters compared to owner-occupiers', *Journal of Social Issues*, 59(3), pp. 527–546. doi: 10.1111/1540-4560.00076.

HM Government (2012) 'No Health Without Mental Health: A cross-government mental health outcomes strategy for people of all ages', *Policy*, pp. 1–103. doi: 10.1016/S0140-6736(07)61238-0.

Hollingshead, A. B. and Redlich, F. C. (1958) 'Social class and mental illness: Community study.', *Social class and mental illness: Community study*. doi: http://dx.doi.org/10.1037/10645-000.

Holt-Lunstad, J., Birmingham, W. and Jones, B. Q. (2008) 'Is there something unique about marriage? The relative impact of marital status, relationship quality, and network social support on ambulatory blood pressure and mental health', *Annals of Behavioral Medicine*, 35(2), pp. 239–244. doi: 10.1007/s12160-008-9018-y.

Horwitz, A. V, White, H. R. and Howell-white, S. (1996) 'Becoming Married and Mental Health : A Longitudinal Study of a Cohort of Young Adults', 58(4), pp. 895–907.

Howell, R. T., Kern, M. L. and Lyubomirsky, S. (2007) 'Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes', *Health Psychology Review*, 1(1), pp. 83–136. doi: 10.1080/17437190701492486.

Hox, J. J. (2002) *Multilevel analysis: Techniques and applications*, *Quantitative methodology series*. doi: 10.1198/jasa.2003.s281.

Hsu, C. Y., Chang, S. Sen and Yip, P. (2017) 'Individual-, household- and neighbourhood-level characteristics associated with life satisfaction: A multilevel analysis of a population-based sample from Hong Kong', *Urban Studies*, 54(16), pp. 3700–3717. doi: 10.1177/0042098016677208.

Hu, L. and Bentler, P. M. (1999) 'Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives', *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), pp. 1–55. doi: 10.1080/10705519909540118.

Hu, Y. *et al.* (2007) 'Can the 12-item General Health Questionnaire be used to measure positive mental health?', *Psychological medicine*, 37(7), pp. 1005–13. doi: 10.1017/S0033291707009993.

Hunt, J. B. *et al.* (2015) 'Racial/ethnic disparities in mental health care utilization among U.S. college students: Applying the institution of medicine definition of health care disparities', *Academic Psychiatry*, 39(5), pp. 520–526. doi: 10.1007/s40596-014-0148-1.

Huppert, F. a (2009) 'Psychological Well-being: Evidence Regarding its Causes and Consequences', *Applied Psychology: Health and Well-Being*, 1(2), pp. 137–164. doi: 10.1111/j.1758-0854.2009.01008.x.

Huppert, F. A. and Whittington, J. E. (2003) 'Evidence for the independence of positive and negative well-being: Implications for quality of life assessment', *British Journal of Health Psychology*, 8(1), pp. 107–122. doi: 10.1348/135910703762879246.

Huppert, F. a, Baylis, N. and Keverne, B. (2005) *Positive Mental Health in Individuals and Populations*, *The Science of Well-being*. doi: 10.1093/acprof.

Jackson-Triche, M. and Sullivan, J. (2000) 'Depression and health-related quality of life in ethnic minorities seeking care in general medical settings.', *Journal of affective disorders*. Available at: http://psycnet.apa.org/psycinfo/2000-03849-001 (Accessed: 23 January 2013).

Jackson, C. (2006) 'The General Health Questionnaire', *Occupational Medicine*, 57(1), pp. 79–79. doi: 10.1093/occmed/kql169.

Jarema, M. and Konieczyńska, Z. (2001) 'Quality of life in schizophrenia: Impact of psychopathology, patients' gender and antipsychotic treatment', *International Journal of Psychiatry in Clinical Practice*, 5(1), pp. 19–26. doi: 10.1080/13651500152048414.

Jelleyman, T. and Spencer, N. (2008) 'Residential mobility in childhood and health outcomes: A systematic review', *Journal of Epidemiology and Community Health*, 62(7), pp. 584–592. doi: 10.1136/jech.2007.060103.

Jones-Webb, R. J. and Snowden, L. R. (1993) 'Symptoms of depression among Blacks and Whites', *American Journal of Public Health*, 83(2), pp. 240–244. doi: 10.2105/AJPH.83.2.240.

Jones, A. M. and Wildman, J. (2008) 'Health, income and relative deprivation: evidence from the BHPS.', *Journal of health economics*, 27(2), pp. 308–24. doi: 10.1016/j.jhealeco.2007.05.007.

Jones, K. (1991) 'Specifying and Estimating Multi-Level Models for Geographical Research', *Transactions of the Institute of British Geographers*, 16(2), pp. 148–159. doi: 10.2307/622610.

Jones, K., Gould, M. I. and Watt, R. (1998) 'Multiple Contexts as Cross-classified Models: The Labor Vote in the British General Election of 1992', *Geographical Analysis*, 30(1), pp. 65–93. doi: 10.1111/j.1538-4632.1998.tb00389.x.

Jones, K., Johnston, R. J. and Pattie, C. J. (1992) 'People, Places and Regions: Exploring the Use of Multi-Level Modelling in the Analysis of Electoral Data', *British Journal of Political Science*, 22(3), p. 343. doi: 10.1017/S0007123400006426.

Jones, K., Johnston, R. and Pattie, C. (1992) 'People, places and regions: exploring the use of multi-level modelling in the analysis of electoral data', *British Journal of Political …*, 22(3), pp. 343–380. Available at: http://journals.cambridge.org/abstract_S0007123400006426 (Accessed: 21 March 2013).

Jöreskog, K. G. (1967) 'A general apporach to confirmatory maximum likelihood factor analysis', *ETS Research Bulletin Series*, 1967(2), pp. 183–202. doi: 10.1002/j.2333-8504.1967.tb00991.x.

Jorm, a F. *et al.* (2005) 'Age group differences in psychological distress: the role of psychosocial risk factors that vary with age.', *Psychological medicine*, 35(9), pp. 1253–63. doi: 10.1017/S0033291705004976.

Kahneman, D. (1999) 'Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), Well-being: The foundations of hedonic psychology', *New York: Russell Sage Foundation.*, pp. 3–25.

Kalliath, T. J., O'Driscoll, M. P. and Brough, P. (2004) 'A confirmatory factor analysis of the General Health Questionnaire-12', *Stress and Health*, 20(1), pp. 11–20. doi: 10.1002/smi.993.

Karlsen, S. *et al.* (2005) 'Racism, psychosis and common mental disorder among ethnic minority groups in England', *Psychological Medicine*, 35(12), pp. 1795–1803. doi: 10.1017/S0033291705005830.

Karlsen, S. and Nazroo, J. Y. (2002) 'Relation between racial discrimination, social class, and health among ethnic minority groups', *American Journal of Public Health*, 92(4), pp. 624–631. doi: 10.2105/AJPH.92.4.624.

Katikireddi, S. V., Niedzwiedz, C. L. and Popham, F. (2012) 'Trends in population mental health before and after the 2008 recession: a repeat cross-sectional analysis of the 1991–2010 Health Surveys of England', *BMJ Open*, 2(5), pp. 1–9. doi: 10.1136/bmjopen-2012-001790.

Keller, M. B. *et al.* (2000) 'A Comparison of Nefazodone, the Cognitive Behavioral-Analysis System of Psychotherapy, and Their Combination for the Treatment of Chronic Depression', *New England Journal of Medicine*, 342(20), pp. 1462–1470. doi: 10.1056/NEJM200005183422001.

Kendler, K. *et al.* (2005) 'The Interaction of Stressful Life Events and a Serotonin Transporter Polymorphism in the Prediction …', *Arch Gen Psychiatry*, 62(May). Available at: http://archpsyc.ama-assn.org/cgi/content/abstract/62/5/529.

Kentikelenis, A. *et al.* (2011) 'Health effects of financial crisis: Omens of a Greek tragedy', *The Lancet*, 378(9801), pp. 1457–1458. doi: 10.1016/S0140-6736(11)61556-0.

Keyes, C. L. M. (2002) 'The Mental Health Continuum: From Languishing to Flourishing in Life', *Journal of Health and Social Behavior*, 43(2), p. 207. doi: 10.2307/3090197.

Keyes, C. L. M. (2005) 'Mental Illness and / or Mental Health ? Investigating Axioms of the Complete State Model of Health', 73(3), pp. 539–548. doi: 10.1037/0022-006X.73.3.539.

Keyes, C. L. M., Dhingra, S. S. and Simoes, E. J. (2010) 'Change in level of positive mental health as a predictor of future risk of mental Illness', *American Journal of Public Health*, 100(12), pp. 2366–2371. doi: 10.2105/AJPH.2010.192245.

Keyes, C. L. M., Shmotkin, D. and Ryff, C. D. (2002) 'Optimizing well-being: the empirical encounter of two traditions.', *Journal of personality and social psychology*, 82(6), pp. 1007–1022. doi: 10.1037/0022-3514.82.6.1007.

Keyes, C. L. M. and Simoes, E. J. (2012) 'To flourish or not: Positive mental health and all-cause mortality', *American Journal of Public Health*, 102(11), pp. 2164–2172. doi: 10.2105/AJPH.2012.300918.

Kiliç, C. *et al.* (1997) 'General Health Questionnaire (GHQ12 and GHQ28): Psychometric properties and factor structure of the scales in a Turkish primary care sample', *Social Psychiatry and Psychiatric Epidemiology*, 32(6), pp. 327–331. doi: 10.1007/BF00805437.

Kilpatrick, D. G. *et al.* (2000) 'Risk factors for adolescent substance abuse and dependence: Data from a national sample.', *Journal of Consulting and Clinical Psychology*, 68(1), pp. 19–30. doi: 10.1037//0022-006X.68.1.19.

Kim, H. K. and McKenry, P. C. (2002) 'The relationship between marriage and psychological well-being: A longitudinal analysis', *Journal of Family Issues*, 23(8), pp. 885–911. doi: 10.1177/019251302237296.

Knapp, M., McDaid, D. and Parsonage, M. (2011) 'Mental health promotion and mental illness prevention: the economic case', *Political Science*, (January), p. 48. Available at: http://www2.lse.ac.uk/LSEHealthAndSocialCare/PSSRU/Home.aspx.

Knies, G. (2017) *Understanding Society: Wave 1-7, 2009-2016 and harmonised British Household Panel Survey: Waves 1-18, 1991-2009, User Guide., Colchester. University of Essex.* doi: 10.2307/3348243.

Lamers, S. M. A. *et al.* (2015) 'The bidirectional relation between positive mental health and psychopathology in a longitudinal representative panel study', *Journal of Positive Psychology*. Routledge, 10(6), pp. 553–560. doi: 10.1080/17439760.2015.1015156.

Lang, I. a *et al.* (2011) 'Income and the midlife peak in common mental disorder prevalence.', *Psychological medicine*, 41(7), pp. 1365–72. doi: 10.1017/S0033291710002060.

Law, R. W. and Sbarra, D. a (2009) 'The effects of church attendance and marital status on the longitudinal trajectories of depressed mood among older adults.', *Journal of aging and health*, 21(6), pp. 803–23. doi: 10.1177/0898264309338300.

Lawlor, C. *et al.* (2012) 'Ethnic variations in pathways to acute care and compulsory detention for women experiencing a mental health crisis', *International Journal of Social Psychiatry*, 58(1), pp. 3–15. doi: 10.1177/0020764010382369.

Layard, R. (2006) 'Happiness and public policy: A challenge to the profession', *Economic Journal*, 116(510). doi: 10.1111/j.1468-0297.2006.01073.x.

Leamy, M. *et al.* (2011) 'Conceptual framework for personal recovery in mental health: Systematic review and narrative synthesis', *British Journal of Psychiatry*, 199(6), pp. 445–452. doi: 10.1192/bjp.bp.110.083733.

Lecrubier, Y. (2007) 'Widespread underrecognition and undertreatment of anxiety and mood disorders: Results from 3 European studies', *Journal of Clinical Psychiatry*, 68(SUPPL. 2), pp. 36–41.

Lépine, J. P. and Briley, M. (2011) 'The increasing burden of depression', *Neuropsychiatric Disease and Treatment*, 7(SUPPL.), pp. 3–7. doi: 10.2147/NDT.S19617.

Leventhal, T. and Brooks-Gunn, J. (2003) 'Moving to opportunity: an experimental study of neighborhood effects on mental health.', *American journal of public health*, 93(9), pp. 1576–82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1448013&tool=pmcentrez&rendertype=abstract.

Levin, K. *et al.* (2012) 'Subjective health and mental well-being of adolescents and the health promoting school: A cross-sectional multilevel analysis', *Health Education*, 112(2), pp. 170–184. doi: 10.1108/09654281211203439.

Levinson, D. *et al.* (2010) 'Associations of serious mental illness with earnings: results from the WHO World Mental Health surveys.', *The British journal of psychiatry : the journal of mental science*, 197(2), pp. 114–21. doi: 10.1192/bjp.bp.109.073635.

Lewis, G. and Booth, M. (1992) 'Regional differences in mental health in Great Britain.', *Journal of epidemiology and community health*, 46(6), pp. 608–11. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1059678&tool=pmcentrez&rendertype=abstract.

Lindstrom, M. and Rosvall, M. (2012) 'Marital status, social capital, economic stress, and mental health: A population-based study.', *The Social Science Journal*. Western Social Science Association, 49(3), pp. 339–342. doi: 10.1016/j.soscij.2012.03.004.

Lorant, V. *et al.* (2003) 'Socioeconomic inequalities in depression: A meta-analysis', *American Journal of Epidemiology*, 157(2), pp. 98–112. doi: 10.1093/aje/kwf182.

Lorant, V. *et al.* (2014) 'Depression and socio-economic risk factors : 7-year longitudinal population study AUTHOR ' S PROOF Depression and socio-economic risk factors : 7-year longitudinal population study', pp. 293–298. doi: 10.1192/bjp.bp.105.020040.

Maas, C. J. M. and Hox, J. J. (2004) ‘Robustness issues in multilevel regression analysis’, *Statistica Neerlandica*, 58(2), pp. 127–137. doi: 10.1046/j.0039-0402.2003.00252.x.

Maheswaran, H. *et al.* (2012) ‘Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): group and individual level analysis.’, *Health and quality of life outcomes*. Health and Quality of Life Outcomes, 10(1), p. 156. doi: 10.1186/1477-7525-10-156.

Mair, C., Diez Roux, a V and Galea, S. (2008) ‘Are neighbourhood characteristics associated with depressive symptoms? A review of evidence.’, *Journal of epidemiology and community health*, 62(11), p. 940–6, 8 p following 946. doi: 10.1136/jech.2007.066605.

Mangalore, R., Knapp, M. and Jenkins, R. (2007) ‘Income-related inequality in mental health in Britain: The concentration index approach’, *Psychological Medicine*, 37(7), pp. 1037–1045. doi: 10.1017/S003329170600969X.

Manley, D., Flowerdew, R. and Steel, D. (2006) ‘Scales, levels and processes: Studying spatial patterns of British census variables’, *Computers, Environment and Urban Systems*, 30(2), pp. 143–160. doi: 10.1016/j.compenvurbsys.2005.08.005.

Marsh, H. W. (2007) ‘Application of confirmatory factor analysis and structural equation modeling in sport and exercise psychology’, in *Handbook of sport psychology*, pp. 774–798.

Marsh, H. W. *et al.* (2009) *Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students’ Evaluations of University Teaching*, *Structural Equation Modeling: A Multidisciplinary Journal*. doi: 10.1080/10705510903008220.

Marsh, H. W. *et al.* (2010) ‘A new look at the big five factor structure through exploratory structural equation modeling.’, *Psychological assessment*, 22(3), pp. 471–491. doi: 10.1037/a0019227.

Marsh, H. W. *et al.* (2014) ‘Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis.’, *Annual review of clinical psychology*, 10(Mimic), pp. 85–110. doi: 10.1146/annurev-clinpsy-032813-153700.

Marsh, H. W. and Bailey, M. (1991) ‘Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models’, *Applied Psychological Measurement*, 15(1), pp. 47–70. doi: 10.1177/014662169101500106.

Marsh, H. W., Hau, K.-T. and Wen, Z. (2004) ‘In Search of Golden Rule : Comment on Hypothesis Testing Approaches to Setting Cutoff Value for Fit Indexes and Danger in Overgeneralizing Hu and Bentler’s (1999) Finding’, *Structural Equation Modeling*, 11(3), pp. 320–341. doi: 10.1207/s15328007sem1103.

Martin, A. J. (1999) ‘Assessing the Multidimensionality of the 12-Item General Health Questionnaire’, *Psychological Reports*, 84, pp. 927–935.

Martin, C. R. and Newell, R. J. (2005) ‘The factor structure of the 12-item General Health Questionnaire in individuals with facial disfigurement.’, *Journal of psychosomatic research*, 59(4), pp. 193–9. doi: 10.1016/j.jpsychores.2005.02.020.

Mason, K. E. *et al.* (2013) ‘Housing affordability and mental health: Does the relationship differ for renters and home purchasers?', *Social Science and Medicine*. Elsevier Ltd, 94, pp. 91–97. doi: 10.1016/j.socscimed.2013.06.023.

Massé, R. *et al.* (1998) 'The structure of mental health: Higher-order confirmatory factor analyses of psychological distress and well-being measures', *Social Indicators Research*, 45, pp. 475–504. doi: 10.1023/A:1006992032387.

Mater, A. *et al.* (2013) 'Measurement Invariance of the Warwick-Edinburgh Mental Well-Being Scale and its Short form across UK and Italy', pp. 1–58.

Mays, V. M. and Cochran, S. D. (2001) 'Mental health correlates of perceived discrimination among lesbian, gay, and bisexual adults in the United States.', *American journal of public health*, 91(11), pp. 1869–76. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1446893&tool=pmcentrez&rendertype=abstract.

McCulloch, C. E. and Neuhaus, J. M. (2011) 'Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter', *Statistical Science*, 26(3), pp. 388–402. doi: 10.1214/11-STS361.

McFall, S. L. (2011) *Understanding Society - The UK household longitudinal study: Wave 1, 2009-2010, User Manual.* Colchester. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Understanding+Society#3 (Accessed: 2 November 2012).

McGuire, T. G. and Miranda, J. (2008) 'New evidence regarding racial and ethnic disparities in mental health: Policy implications', *Health Affairs*, 27(2), pp. 393–403. doi: 10.1377/hlthaff.27.2.393.

McKee-Ryan, F. *et al.* (2005) 'Psychological and physical well-being during unemployment: a meta-analytic study.', *The Journal of applied psychology*, 90(1), pp. 53–76. doi: 10.1037/0021-9010.90.1.53.

McLean, C. P. *et al.* (2011) 'Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness.', *Journal of psychiatric research*. Elsevier Ltd, 45(8), pp. 1027–35. doi: 10.1016/j.jpsychires.2011.03.006.

McManus, S. *et al.* (2016) 'Mental health and well-being in England: Adult Psychiatric Morbidity Survey 2014'.

Meltzer, H. *et al.* (2012) 'The relationship between personal debt and specific common mental disorders.', *European journal of public health*, pp. 1–6. doi: 10.1093/eurpub/cks021.

Memetovic, J., Ratner, P. A. and Richardson, C. G. (2014) 'Gender-based measurement invariance of the substance use risk profile scale', *Addictive Behaviors*. Elsevier Ltd, 39(3), pp. 690–694. doi: 10.1016/j.addbeh.2013.10.016.

Mezuk, B., Myers, J. M. and Kendler, K. S. (2013) 'Integrating social science and behavioral genetics: testing the origin of socioeconomic disparities in depression using a genetically informed design.', *American journal of public health*, 103 Suppl, pp. S145-51. doi: 10.2105/AJPH.2013.301247.

Milne, E. (2014) 'North and South: addressing the English health divide', *Journal of public health (Oxford, England)*, 36(2), pp. 183–186. doi: 10.1093/pubmed/fdu029.

Moerbeek, M. (2004) 'The consequence of ignoring a level of nesting in multilevel analysis', *Multivariate Behavioral Research*, (May 2014), pp. 37–41. doi: 10.1207/s15327906mbr3901.

Moller, H. *et al.* (2013) 'Rising unemployment and increasing spatial health inequalities in England: Further extension of the North-South divide', *Journal of Public Health (United Kingdom)*, 35(2), pp. 313–321. doi: 10.1093/pubmed/fds085.

Morgan, C. *et al.* (2009) 'Ethnicity, social disadvantage and psychotic-like experiences in a healthy population based sample', *Acta Psychiatrica Scandinavica*, 119(3), pp. 226–235. doi: 10.1111/j.1600-0447.2008.01301.x.

Morris, T. *et al.* (2017) 'How do moving and other major life events impact mental health? A longitudinal analysis of UK children', *Health and Place*. Elsevier Ltd, 46(January), pp. 257–266. doi: 10.1016/j.healthplace.2017.06.004.

Mukuria, C. *et al.* (2014) 'An empirical comparison of well-being measures used in UK', (September), pp. 1–55.

Muntaner, C. *et al.* (2004) 'Socioeconomic position and major mental disorders.', *Epidemiologic reviews*, 26, pp. 53–62. doi: 10.1093/epirev/mxh001.

Murphy, G. C. and Athanasou, J. A. (1999) 'The effect of unemployment on mental health', *Journal of Occupational and Organizational Psychology*, 72(1), pp. 83–99. doi: 10.1348/096317999166518.

Murray, G. *et al.* (2008) 'Big boys don't cry: An investigation of stoicism and its mental health outcomes', *Personality and Individual Differences*, 44(6), pp. 1369–1381. doi: 10.1016/j.paid.2007.12.005.

Muth, B. (2010) 'Bayesian Analysis In Mplus : A Brief Introduction'.

Muthén, B. and Asparouhov, T. (2012) 'Bayesian structural equation modeling: a more flexible representation of substantive theory.', *Psychological methods*, 17(3), pp. 313–35. doi: 10.1037/a0026802.

Neeleman, J., Ormel, J. and Bijl, R. V (2001) 'The distribution of psychiatric and somatic Ill health: associations with personality and socioeconomic status', *Psychosomatic medicine*, 63, pp. 239–247. doi: 10.1097/00006842-200103000-00007.

van der Noordt, M. *et al.* (2014) 'Health effects of employment: a systematic review of prospective studies.', *Occupational and environmental medicine*, 71(10), pp. 730–6. doi: 10.1136/oemed-2013-101891.

Oguz, S., Merad, S. and Snape, D. (2013) 'Measuring National Well-being - What matters most to Personal Well-being ?', *Report*, (May), pp. 1–59.

Oppenheimer, V. K. (2000) 'The Continuing Importance of Men's Economic Position in Marriage Formation', in *The Ties That Bind: Perspectives on Marriage and Cohabitation*, pp. 283–301. Available at: http://books.google.com/books?hl=en&lr=&id=lEb-

ic22xKwC&pgis=1.

Van Os, J., Park, S. B. G. and Jones, P. B. (2001) 'Neuroticism, life events and mental health: Evidence for person-environment correlation', *British Journal of Psychiatry*, 178(SUPPL. 40). doi: 10.1192/bjp.178.40.s72.

Padrón, A. *et al.* (2012) 'Confirmatory factor analysis of the General Health Questionnaire (GHQ-12) in Spanish adolescents.', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 21(7), pp. 1291–8. doi: 10.1007/s11136-011-0038-x.

Park, S. *et al.* (2017) 'Lay beliefs about the causes and cures of schizophrenia', *International Journal of Social Psychiatry*, 63(6), pp. 518–524. doi: 10.1177/0020764017717283.

Pattyn, E., Verhaeghe, M. and Bracke, P. (2015) 'The gender gap in mental health service use', *Social Psychiatry and Psychiatric Epidemiology*. Springer Berlin Heidelberg, 50(7), pp. 1089–1095. doi: 10.1007/s00127-015-1038-x.

Paul, K. I. and Moser, K. (2009) 'Unemployment impairs mental health: Meta-analyses', *Journal of Vocational Behavior*. Elsevier Inc., 74(3), pp. 264–282. doi: 10.1016/j.jvb.2009.01.001.

Penninkilampi-Kerola, V., Miettunen, J. and Ebeling, H. (2006) 'A comparative assessment of the factor structures and psychometric properties of the GHQ-12 and the GHQ-20 based on data from a Finnish population-based sample.', *Scandinavian journal of psychology*, 47(5), pp. 431–40. doi: 10.1111/j.1467-9450.2006.00551.x.

Petrosky, M. J. and Birkimer, J. C. (1991) 'The relationship among locus of control, coping styles, and psychological symptom reporting', *Journal of Clinical Psychology*, 47(3), pp. 336–345. doi: 10.1002/1097-4679(199105)47:3<336::AID-JCLP2270470303>3.0.CO;2-L.

Pevalin, D. J. (2000) 'Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest effects', *Social Psychiatry and Psychiatric Epidemiology*, 35(11), pp. 508–512. doi: 10.1007/s001270050272.

Phillips, D. L., Andrews, F. M. and Withey, S. B. (1978) 'Social Indicators of Well-Being: Americans' Perceptions of Life Quality.', *Contemporary Sociology*, p. 392. doi: 10.2307/2064314.

Picardi, a, Abeni, D. and Pasquini, P. (2001) 'Assessing psychological distress in patients with skin diseases: reliability, validity and factor structure of the GHQ-12', *Journal of the European Academy of Dermatology and Venereology*, 15(5), pp. 410–417. doi: 10.1046/j.1468-3083.2001.00336.x.

Pickett, K. E. and Pearl, M. (2001) 'Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review.', *Journal of epidemiology and community health*, 55(2), pp. 111–22. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1731829&tool=pmcentrez&rendertype=abstract.

Pierse, N. *et al.* (2016) 'Examining the role of tenure, household crowding and housing affordability on psychological distress, using longitudinal data', *Journal of Epidemiology and*

Community Health, pp. 961–966. doi: 10.1136/jech-2015-206716.

Politi, P. L., Piccinelli, M. and Wilkinson, G. (1994) 'Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy.', *Acta psychiatrica Scandinavica*, 90(6), pp. 432–437. doi: 10.1111/j.1600-0447.1994.tb01620.x.

Porter, M. and Haslam, N. (2005) 'Predisplacement and postdisplacement of refugees and internally displaced persons', *The Journal of the American Medical Association*, 294(5), pp. 610–612. doi: 10.1001/jama.294.5.602.

Propper, C. *et al.* (2005) 'Local neighbourhood and mental health: evidence from the UK.', *Social science & medicine (1982)*, 61(10), pp. 2065–83. doi: 10.1016/j.socscimed.2005.04.013.

Puustinen, P. J. *et al.* (2011) 'Psychological distress measured by the GHQ-12 and mortality: a prospective population-based study.', *Scandinavian journal of public health*, 39(6), pp. 577–81. doi: 10.1177/1403494811414244.

Rahmati Najarkolaei, F. *et al.* (2014) 'Factor Structure of the Iranian Version of 12-Item General Health Questionnaire', *Iranian Red Crescent Medical Journal*, 16(9). doi: 10.5812/ircmj.11794.

Randall, C. (Office for N. S. (2012) 'Measuring National Well-being: Where We Live, 2012', (September), pp. 1–42.

Rasbash, J. and Goldstein, H. (1994) 'Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model', *Journal of Educational and Behavioral Statistics*, 19(4), pp. 337–350. doi: 10.3102/10769986019004337.

Reeves, A. *et al.* (2012) 'Increase in state suicide rates in the USA during economic recession', *The Lancet*. Elsevier Ltd, 380(9856), pp. 1813–1814. doi: 10.1016/S0140-6736(12)61910-2.

Reeves, A., McKee, M. and Stuckler, D. (2014) 'Economic suicides in the Great Recession in Europe and North America', *British Journal of Psychiatry*, 205(3), pp. 246–247. doi: 10.1192/bjp.bp.114.144766.

Ritsher, J. E. B. (2001) 'Inter-generational longitudinal study of social class and depression: a test of social causation and social selection models', *The British Journal of Psychiatry*, 178(40), p. 84s–90. doi: 10.1192/bjp.178.40.s84.

Romppel, M. *et al.* (2013) 'What is the General Health Questionnaire-12 assessing?: Dimensionality and psychometric properties of the General Health Questionnaire-12 in a large scale German population sample', *Comprehensive Psychiatry*. Elsevier Inc., 54(4), pp. 406–413. doi: 10.1016/j.comppsych.2012.10.010.

Ross, C. (2000) 'Neighborhood disadvantage and adult depression', *Journal of Health and Social Behavior*, 41(2), pp. 177–187. Available at: http://www.jstor.org/stable/10.2307/2676304 (Accessed: 4 November 2012).

Ross, C. E., Reynolds, J. R. and Geis, K. J. (2000) 'The contingent meaning of neighbourhood stability for residents' psychological well-being.', *American Sociological Review*, 65(4), pp. 581–597.

Rothman, K. J., Gallacher, J. E. J. and Hatch, E. E. (2013) 'Why representativeness should be avoided', *International Journal of Epidemiology*, 42(4), pp. 1012–1014. doi: 10.1093/ije/dys223.

Rubin, D. B. (1976) 'Inference and missing Data - Reply', *Biometrika*, 63(3), pp. 591–592.

Ryan, R. M. and Deci, E. L. (2001) 'ON HAPPINESS AND HUMAN POTENTIALS: A Review of Research on Hedonic and Eudaimonic Well-Being', *Annu Rev Psychol*, 52, pp. 141–166.

Ryff, C. D. (2013) 'Psychological well-being revisited: Advances in the science and practice of eudaimonia', *Psychotherapy and Psychosomatics*, 83(1), pp. 10–28. doi: 10.1159/000353263.

Samaritans (2015) 'Suicide Statistics Report 2015'.

Sánchez-López, M. D. P. and Dresch, V. (2008) 'The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population.', *Psicothema*, 20(4), pp. 839–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18940092.

Sass, D. A. and Schmitt, T. A. (2010) 'A Comparative Investigation of Rotation Criteria Within Exploratory Factor Analysis', *Multivariate Behavioral Research*, 45(1), pp. 73–103. doi: 10.1080/00273170903504810.

Schmitt, T. a. (2011) 'Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis', *Journal of Psychoeducational Assessment*, 29(4), pp. 304–321. doi: 10.1177/0734282911406653.

Schmitz, N., Kruse, J. and Tress, W. (1999) 'Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample.', *Acta psychiatrica Scandinavica*, 100(6), pp. 462–468. doi: 10.1111/j.1600-0447.1999.tb10898.x.

Schomerus, G. *et al.* (2012) 'Evolution of public attitudes about mental illness: a systematic review and meta-analysis.', *Acta psychiatrica Scandinavica*, 125(6), pp. 440–52. doi: 10.1111/j.1600-0447.2012.01826.x.

Schrank, B. *et al.* (2013) 'Conceptualising and measuring the well-being of people with psychosis: Systematic review and narrative synthesis', *Social Science and Medicine*. Elsevier Ltd, 92, pp. 9–21. doi: 10.1016/j.socscimed.2013.05.011.

Self, S. and Liang, K. (1987) 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions.', *J Am Stat Assoc*, 82(398), pp. 605–610. doi: 10.1080/01621459.1987.10478472.

Shevlin, M. and Adamson, G. (2005) 'Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis.', *Psychological assessment*, 17(2), pp. 231–236. doi: 10.1037/1040-3590.17.2.231.

Silver, E., Mulvey, E. P. and Swanson, J. W. (2002) 'Neighborhood structural characteristics and mental disorder: Faris and Dunham revisited.', *Social science & medicine (1982)*, 55(8), pp. 1457–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12231022.

Singh, S. P. *et al.* (2012) 'Ethnicity and the Mental Health Act 1983.', *Health (San Francisco)*, pp. 99–105. doi: 10.1192/bjp.bp.106.030346.

Singh, S. P. *et al.* (2014) 'Ethnicity as a predictor of detention under the Mental Health Act', *Psychological Medicine*, 44(5), pp. 997–1004. doi: 10.1017/S003329171300086X.

Slade, M. (2010) 'Mental illness and well-being: the central importance of positive psychology and recovery approaches.', *BMC health services research*, 10, p. 26. doi: 10.1186/1472-6963-10-26.

Smith, A. B. *et al.* (2010) 'A Rasch and confirmatory factor analysis of the general health questionnaire (GHQ)--12.', *Health and quality of life outcomes*, 8, p. 45. doi: 10.1186/1477-7525-8-45.

Smith, A. B. *et al.* (2013) 'The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress.', *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 22(1), pp. 145–52. doi: 10.1007/s11136-012-0133-7.

Solnick, S. J. and Hemenway, D. (1998) 'Is more always better?: A survey on positional concerns', *Journal of Economic Behavior & Organization*, 37, pp. 373–383. doi: 10.1016/S0167-2681(98)00089-4.

Spence, R. *et al.* (2014) 'Focus On: Antidepressant prescribing. Trends in the prescribing of antidepressants in primary care.', *The Health Foundation and the Nuffield Trust*, (May), pp. 1–39. Available at: http://www.nuffieldtrust.org.uk/publications/focus-antidepressant-prescribing.

Spiegelhalter, D. *et al.* (2002) 'Bayesian measures of model complexity and fit', *Journal of the Royal ...*, pp. 583–639. Available at: http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00353/full (Accessed: 12 May 2014).

Steele, F. (2008) 'Multilevel models for longitudinal data', *Journal of the Royal Statistical Society. Series A*, 171(1), pp. 5–19. doi: 10.1111/j.1467-985X.2007.00509.x.

Steptoe, A., Deaton, A. and Stone, A. A. (2015) 'Subjective wellbeing, health, and ageing', *The Lancet*. Elsevier Ltd, 385(9968), pp. 640–648. doi: 10.1016/S0140-6736(13)61489-0.

Stewart-Brown, S. *et al.* (2009) 'Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey.', *Health and quality of life outcomes*, 7, p. 15. doi: 10.1186/1477-7525-7-15.

Stewart-Brown, S., Samaraweera, P. C., Taggart, F., Kandala, N. B., *et al.* (2015) 'Socioeconomic gradients and mental health: Implications for public health', *British Journal of Psychiatry*, 206(6), pp. 461–465. doi: 10.1192/bjp.bp.114.147280.

Stewart-Brown, S., Samaraweera, P. C., Taggart, F., Kandala, N.-B., *et al.* (2015) 'Socioeconomic gradients and mental health: implications for public health.', *The British journal of psychiatry : the journal of mental science*, 206(6), pp. 461–5. doi: 10.1192/bjp.bp.114.147280.

Stewart-Brown, S. L. *et al.* (2011) ‘Plenary session : the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) : A valid and reliable tool for measuring mental well-being in diverse populations and projects’, *Journal of Epidemiology and Community Health*, 65(Suppl II), pp. 38–39. doi: 10.1136/jech.2011.143586.86.

Stiglitz, J. E., Sen, A. and Fitoussi, J.-P. (2009) ‘The Measurement of Economic Performance and Social Progress Revisited - Reflections and Overview’, p. 79. Available at: http://www.stiglitz-sen-fitoussi.fr/en/documents.htm.

Stiglitz, J., Sen, A. and Fitoussi, J.-P. (2010) ‘Report by the Commission on the Measurement of Economic Performance and Social Progress (2009)’, *SSRN Electronic Journal*. doi: 10.2139/ssrn.1714428.

Strandh, M. (2000) ‘Different exit routes from unemployment and their impact on mental well-being’, *Work, Employment and Society*, 14(3), pp. 459–479. Available at: http://wes.sagepub.com/content/14/3/459.full.pdf.

Stuckler, D. *et al.* (2009) ‘The public health effect of economic crises and alternative policy responses in Europe: an empirical analysis’, *The Lancet*. Elsevier Ltd, 374(9686), pp. 315–323. doi: 10.1016/S0140-6736(09)61124-7.

Stuckler, D. *et al.* (2011) ‘Effects of the 2008 recession on health: A first look at European data’, *The Lancet*, 378(9786), pp. 124–125. doi: 10.1016/S0140-6736(11)61079-9.

Stutzer, A. *et al.* (2004) ‘The Role of Income Aspirations in Individual Happiness The Role of Income Aspirations in Individual Happiness The Role of Income Aspirations in Individual Happiness’, *Journal of Economic Behavior and Organization*, 54(1), pp. 89–109. doi: 10.1016/j.jebo.2003.04.003.

Subramanian, S. V, Kim, D. and Kawachi, I. (2005) ‘Covariation in the socioeconomic determinants of self rated health and happiness: a multivariate multilevel analysis of individuals and communities in the USA.’, *Journal of epidemiology and community health*, 59(8), pp. 664–9. doi: 10.1136/jech.2004.025742.

Sugiyama, T. *et al.* (2008) ‘Associations of neighbourhood greenness with physical and mental health: do walking, social coherence and local social interaction explain the relationships?’, *Journal of Epidemiology & Community Health*, 62(5), pp. e9–e9. doi: 10.1136/jech.2007.064287.

Suglia, S. F., Duarte, C. S. and Sandel, M. T. (2011) ‘Housing quality, housing instability, and maternal mental health’, *Journal of Urban Health*, 88(6), pp. 1105–1116. doi: 10.1007/s11524-011-9587-0.

Suzuki, H. *et al.* (2011) ‘Clarification of the factor structure of the 12-item General Health Questionnaire among Japanese adolescents and associated sleep status.’, *Psychiatry research*. Elsevier Ltd, 188(1), pp. 138–46. doi: 10.1016/j.psychres.2010.10.025.

Swanson, J. *et al.* (2009) ‘Racial disparities in involuntary outpatient commitment: Are they real?’, *Health Affairs*, 28(3), pp. 816–826. doi: 10.1377/hlthaff.28.3.816.

Taggart, F. *et al.* (2013) ‘Cross cultural evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) --a mixed methods study.’, *Health and quality of life outcomes*.

Health and Quality of Life Outcomes, 11(1), p. 27. doi: 10.1186/1477-7525-11-27.

Tait, R. J., French, D. J. and Hulse, G. K. (2003) 'Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents', *Australian and New Zealand Journal of Psychiatry*, 37(3), pp. 374–381. doi: 10.1046/j.1440-1614.2003.01133.x.

Taris, T. W. (2002) 'Unemployment and mental health: A longitudinal perspective', *International Journal of Stress Management*, 9(1), pp. 43–57. doi: 10.1023/A:1013067101217.

Taylor, M. P., Pevalin, D. J. and Todd, J. (2007) 'The psychological costs of unsustainable housing commitments.', *Psychological medicine*, 37(7), pp. 1027–36. doi: 10.1017/S0033291706009767.

Di Tella, R. and MacCulloch, R. (2008) 'Gross national happiness as an answer to the Easterlin Paradox?', *Journal of Development Economics*, 86(1), pp. 22–42. doi: 10.1016/j.jdeveco.2007.06.008.

Tennant, R. *et al.* (2007) 'The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation.', *Health and quality of life outcomes*, 5, p. 63. doi: 10.1186/1477-7525-5-63.

The Scottish Government (2009) *Towards a mentally flourishing Scotland: Policy and action plan 2009-2011*.

Thomas, C., Benzeval, M. and Stansfeld, S. a (2005) 'Employment transitions and mental health: an analysis from the British household panel survey.', *Journal of epidemiology and community health*, 59(3), pp. 243–9. doi: 10.1136/jech.2004.019778.

Thornicroft, G. (2007) 'Most people with mental illness are not treated', *The Lancet*, 370, pp. 807–808. doi: 10.1016/S0140-6736(07)61242-2.Depression.

Tomaszewski, W. and Perales, F. (2014) 'Who settles for less? Subjective dispositions, objective circumstances, and housing satisfaction', *Social Indicators Research*, 118(1), pp. 181–203. doi: 10.1007/s11205-013-0420-x.

Tomlinson, M. *et al.* (2009) 'Setting priorities for global mental health research', *Bulletin of the World Health Organization*, 87(6), pp. 438–446. doi: 10.2471/BLT.08.054353.

Tranmer, M. and Steel, D. G. (2001) 'Ignoring a level in a multilevel model: evidence from UK census data', *Environment and Planning A*, 33(5), pp. 941–948. doi: 10.1068/a3317.

Trompetter, H. R. *et al.* (2017) 'Both positive mental health and psychopathology should be monitored in psychotherapy: Confirmation for the dual-factor model in acceptance and commitment therapy', *Behaviour Research and Therapy*. Elsevier Ltd, 91, pp. 58–63. doi: 10.1016/j.brat.2017.01.008.

Trompetter, H. R., de Kleine, E. and Bohlmeijer, E. T. (2017) 'Why Does Positive Mental Health Buffer Against Psychopathology? An Exploratory Study on Self-Compassion as a Resilience Mechanism and Adaptive Emotion Regulation Strategy', *Cognitive Therapy and Research*. Springer US, 41(3), pp. 459–468. doi: 10.1007/s10608-016-9774-0.

Uutela, A. (2010) 'Economic crisis and mental health.', *Current opinion in psychiatry*, 23, pp.

127–130. doi: 10.1097/YCO.0b013e328336657d.

Vanheule, S. and Bogaerts, S. (2005) 'The factorial structure of the GHQ-12', *Stress and Health*, 21(4), pp. 217–222. doi: 10.1002/smi.1058.

Verheij, R. (1996) 'Explaining urban-rural variations in health: a review of interactions between individual and environment', *Social science & medicine*, 42(6), pp. 923–935. Available at: http://www.sciencedirect.com/science/article/pii/0277953695001905 (Accessed: 2 November 2012).

Vickers, D. and Rees, P. (2006) 'Introducing the area classification of output areas.', *Population trends*, (125), pp. 15–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17042312.

Vinkers, D. J. *et al.* (2010) 'Ethnicity and dangerousness criteria for court ordered admission to a psychiatric hospital', *Social Psychiatry and Psychiatric Epidemiology*, 45(2), pp. 221–224. doi: 10.1007/s00127-009-0058-9.

De Vogli, R., Marmot, M. and Stuckler, D. (2012) 'Excess suicides and attempted suicides in Italy attributable to the great recession', *Journal of Epidemiology & Community Health*, pp. 1–2. doi: 10.1136/jech-2012-201607.

Wade, T. J. and Pevalin, D. J. (2004) 'Marital Transitions and Mental Health', *Journal of Health and Social Behavior*, 45(2), pp. 155–170. doi: 10.1177/002214650404500203.

Waite, L. J. (2009) 'Marital History and Well-Being in Later Life', *International Handbook of Population Aging*, pp. 691–704. doi: 10.1007/978-1-4020-835.

Wang, P. S. *et al.* (2007) 'Use of mental health services for anxiety, mood, and substance use disorders in 17 countries in the WHO world mental health surveys', *Lancet*, 370(9590), pp. 841–850. doi: 10.1016/S0140-6736(07)61414-7.

Waterman, A. S. (1993) 'Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment.', *Journal of Personality and Social Psychology*, 64(4), pp. 678–691. doi: 10.1037/0022-3514.64.4.678.

Waterman, A. S., Schwartz, S. J. and Conti, R. (2008) 'The implications of two conceptions of happiness (hedonic enjoyment and eudaimonia) for the understanding of intrinsic motivation', *Journal of Happiness Studies*, 9(1), pp. 41–79. doi: 10.1007/s10902-006-9020-7.

Weich, S. (2003) 'Geographic Variation in the Prevalence of Common Mental Disorders in Britain: A Multilevel Investigation', *American Journal of Epidemiology*, 157(8), pp. 730–737. doi: 10.1093/aje/kwg035.

Weich, S. *et al.* (2004) 'Common mental disorders and ethnicity in England', *Psychol Med*, 34(8), pp. 1543–1551. Available at: http://journals.cambridge.org/bin/bladerunner?REQUNIQ=1099652087&REQSESS=2881703&117000REQEVENT=&REQINT1=257758&REQAUTH=0.

Weich, S. *et al.* (2011) 'Mental well-being and mental illness: Findings from the adult psychiatric morbidity survey for England 2007', *British Journal of Psychiatry*, 199(1), pp. 23–28. doi: 10.1192/bjp.bp.111.091496.

Weich, S. and Lewis, G. (1998) 'Material standard of living, social class, and the prevalence of the common mental disorders in Great Britain.', *Journal of epidemiology and community health*, 52(1), pp. 8–14. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1756605&tool=pmcentrez&rendertype=abstract.

Weich, S., Twigg, L. and Lewis, G. (2006) 'Rural/non-rural differences in rates of common mental disorders in Britain: prospective multilevel cohort study.', *The British journal of psychiatry : the journal of mental science*, 188, pp. 51–7. doi: 10.1192/bjp.bp.105.008714.

Werneke, U. *et al.* (2000) 'The stability of the factor structure of the General Health Questionnaire', *Psychological Medicine*, 30(4), pp. 823–829. doi: 10.1017/S0033291799002287.

Westerhof, G. J. and Keyes, Æ. C. L. M. (2010) 'Mental Illness and Mental Health : The Two Continua Model Across the Lifespan', pp. 110–119. doi: 10.1007/s10804-009-9082-y.

Whaley, A. L. (2001) 'Cultural Mistrust and Mental Health Services for African Americans', *The Counseling Psychologist*, 29(4), pp. 513–531. doi: 10.1177/0011000001294003.

Whiteford, H. A. *et al.* (2013) 'Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010', *The Lancet*, 382(9904), pp. 1575–1586. doi: 10.1016/S0140-6736(13)61611-6.

Wikman, A., Wardle, J. and Steptoe, A. (2011) 'Quality of life and affective well-being in middle-aged and older people with chronic medical illnesses: A cross-sectional population based study', *PLoS ONE*, 6(4). doi: 10.1371/journal.pone.0018952.

Williams, D. R. *et al.* (2010) 'Race, socioeconomic status, and health: Complexities, ongoing challenges, and research opportunities', *Annals of the New York Academy of Sciences*, 1186, pp. 69–101. doi: 10.1111/j.1749-6632.2009.05339.x.

Williams, K., Frech, A. and Carlson, D. L. (2010a) 'Marital status and mental health.', in *A handbook for the study of mental health: Social contexts, theories, and systems (2nd ed).*, pp. 306–320. Available at: http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=2010-00304-016&site=ehost-live&scope=site.

Williams, K., Frech, A. and Carlson, D. L. (2010b) 'Marital status and mental health.', in *A handbook for the study of mental health: Social contexts, theories, and systems (2nd ed).*, pp. 306–320.

Williams, K. and Umberson, D. (2004) 'Marital status, marital transitions, and health: A gendered life course perspective', *Journal of Health and Social Behavior*, 45(1), pp. 81–98. doi: 10.1177/002214650404500106.

Wilson, C. and Oswald, A. (2005) 'How does marriage affect physical and psychological health? A survey of the longitudinal evidence', *IZA Discussion Paper No. 1619*, (1619). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=735205 (Accessed: 29 January 2013).

Windle, G. (2011) 'What is resilience? A review and concept analysis', *Reviews in Clinical Gerontology*, 21(2), pp. 152–169. doi: 10.1017/S0959259810000420.

Windle, G. (2012) 'The contribution of resilience to healthy ageing', *Perspectives in Public Health*, 132(4), pp. 159–160. doi: 10.1177/1757913912449572.

Winefield, H. R. *et al.* (1989) 'The General Health Questionnaire: Reliability and validity for Australian youth', *Australian and New Zealand Journal of Psychiatry*, 23(1), pp. 53–58. doi: 10.3109/00048678909062592.

Wittchen, H. U. and Jacobi, F. (2005) 'Size and burden of mental disorders in Europe - A critical review and appraisal of 27 studies', *European Neuropsychopharmacology*, 15(4), pp. 357–376. doi: 10.1016/j.euroneuro.2005.04.012.

World Health Organization (2001) *Strengthening mental health promotion, Fact Sheet No 220.* Available at: http://www.who.int/mediacentre/factsheets/fs220/en/.

World Health Organization (2002) 'Gender and Mental Health', *Department of Gender and Women's Health Department of Mental Health and Substance Dependence*, (June), p. Mental Health. (2002). World Health Organization,. doi: 10.1023/A:1020133308188.

World Health Organization (2004) 'Promoting Mental Health: Concepts, Emerging Evidence, Practice: Summary Report. Geneva, World Health Organization.', *Http://Www.Who.Int/Mental_Health/Evidence/En/Promoting_Mhh.Pdf*, p. 68. doi: 10.1111/j.1466-7657.2004.00268.x.

World Health Organization (2013) 'Mental Health Action Plan 2013-2020', *WHO Library Cataloguing-in-Publication DataLibrary Cataloguing-in-Publication Data*, pp. 1–44. doi: ISBN 978 92 4 150602 1.

Worsley, A. and Gribbin, C. C. (1977) 'A Factor Analytic Study of the Twelve Item General Health Questionnaire', (March), pp. 269–272.

Wurm, S. and Benyamini, Y. (2014) 'Optimism buffers the detrimental effect of negative self-perceptions of ageing on physical and mental health', *Psychology and Health*. Routledge, 29(7), pp. 832–848. doi: 10.1080/08870446.2014.891737.

Ye, S. (2009) 'Factor structure of the General Health Questionnaire (GHQ-12): The role of wording effects', *Personality and Individual Differences*. Elsevier Ltd, 46(2), pp. 197–201. doi: 10.1016/j.paid.2008.09.027.

Yu, C.-Y. and Muthén, B. (2002) 'Evaluating cutoff criteria of model fit indices for latent variable models with binary and continous outcomes.', *Unpublished PhD Thesis*, PhD, p. 168. doi: 10.1.1.310.3956.

Zatzick, D. F. *et al.* (1997) 'Posttraumatic stress disorder and functioning and quality of life outcomes in a nationally representative sample of male Vietnam veterans', *American Journal of Psychiatry*, 154(12), pp. 1690–1695. doi: 10.1176/ajp.154.12.1690.